ESTADÍSTICA MATEMÁTICA CON APLICACIONES

Sexta edición

Prentice Hall

JOHN E. FREUND ▪ IRWIN MILLER ▪ MARYLEES MILLER

Estadística matemática con aplicaciones

Sexta edición

John E. Freund
Arizona State University

Irwin Miller
Marylees Miller

TRADUCCIÓN:
Ing. Rosendo José Sánchez Palma
Ingeniería Química, Instituto Tecnológico y de Estudios Superiores de Monterrey
Maestría en Ciencias, Universidad de Wisconsin

REVISIÓN TÉCNICA:
Carlos Armando Martínez Reyes
Lic. en Física y Matemáticas, Instituto Tecnológico y de Estudios Superiores de Monterrey, Campus Estado de México

MÉXICO • ARGENTINA • BRASIL • COLOMBIA • COSTA RICA • CHILE
ESPAÑA • GUATEMALA • PERÚ • PUERTO RICO • VENEZUELA

Datos de catalogación bibliográfica

Miller, Irwin
Estadística matemática con aplicaciones, 6a. ed.
PEARSON EDUCACIÓN, México, 2000

ISBN: 970-17-0389-8
Área: Universitarios

Formato: 18.5 × 23.5 cm **Páginas: 640**

Versión en español de la obra titulada *John E. Freund's mathematical statistics. Sixth Edition*, de Irwin Miller y Marylees Miller, publicada originalmente en inglés por Prentice Hall Inc., Upper Saddle River, New Jersey, U.S.A.

Esta edición en español es la única autorizada.

Original English language title by
Prentice Hall Inc.

ISBN 0-13-123613-X

Edición en español:
Editor: Guillermo Trujano Mendoza
e-mail: guillermo.trujano@pearsoned.com
Supervisora de traducción: Catalina Pelayo Rojas
Supervisor de producción: Alejandro A. Gómez Ruiz

Edición en inglés:
Acquisition Editor: Ann Heath
Editorial Assistant/Supplement Editor: Mindy McClard
Editorial Director: Tim Bozik
Editor-in-Chief: Jerome Grant
Director of Marketing: John Tweeddale
Marketing Manager: Melody Marcus
Art Director/Cover Designer: Jayne Conte

SEXTA EDICIÓN, 2000

Cámara Nacional de la Industria Editorial Mexicana Reg. Núm. 1031.

ISBN 970-17-0389-8

Impreso en México. *Printed in Mexico.*

1 2 3 4 5 6 7 8 9 0 - 03 02 01 00

Contenido

Prefacio

La sexta edición de *Estadística matemática con aplicaciones*, al igual que las primeras cinco ediciones, está diseñada para un curso, con base en el cálculo de introducción a las matemáticas de la estadística de dos semestres o tres trimestres. Esta edición realza los cambios que se hicieron en la quinta edición para reflejar los cambios que, en años recientes, han tenido lugar en el pensamiento estadístico y en la enseñanza de la estadística.

Se ha puesto más énfasis en el uso de las computadoras al efectuar cálculos estadísticos. Se han añadido varios ejercicios nuevos, muchos de los cuales requieren el uso de una computadora. Además, se ha añadido nuevo material al capítulo 15, entre el que se encuentra el modelo de análisis de la varianza en dos sentidos con interacción y una revisión de las comparaciones múltiples. También, se han añadido los apéndices, que resumen las propiedades de las funciones de distribución y densidad de probabilidad especiales que aparecen en el texto.

Agradecemos tantos comentarios constructivos que recibimos del Dr. John E. Freund y de los siguientes revisores: D. S. Gill, California State Polytechnic University, Pomona; Joseph Walker, Georgia State University; Susan Herring, Sonoma State University; y Geetha Ramachandran, California State University, Sacramento. También queremos hacer un reconocimiento a las valiosas contribuciones del difunto Dr. Ronald E. Walpole a la tercera y cuarta ediciones.

Asimismo deseamos expresar nuestro aprecio a Robert E. Krieger Publishing Company por la autorización para basar la tabla II en la obra *Poisson's Exponential Binomial Limit* de E. C. Molina; a Prentice Hall, Inc. por la autorización para reproducir parte de la tabla IV de *Applied Multivariate Statistical Analysis* de R. A. Johnson y D. W. Wichern; al profesor E. S. Pearson y a los fiduciarios de *Biometrika* por la reproducción del material en las tablas V y VI; a los editores de *Biometrics* por la autorización para reproducir el material de "Critical Values for Duncan's New Multiple Range Test" de H. L. Harter para la tabla IX; a American Cyanamid Company por la reproducción del material de *Some Rapid Approximate Statistical Procedures* de F. Wilcoxon y R. A. Wilcox para la tabla X; a D. Auble por la fundamentación

de la tabla XI en su "Extended Tables for the Mann-Whitney Statistics," *Bulletin of the Institute of Educational Research at Indiana University*; al editor de *Annals of Mathematical Statistics* por la reproducción del material en la tabla XII; y a MINITAB® por la reproducción de las salidas impresas de computadora que se muestran en el texto.

A los autores también les gustaría expresar su aprecio al personal de Prentice Hall, en especial a Elaine Wetterau, por su atenta cooperación para la producción de este libro.

Irwin Miller
Marylees Miller

Hampton, New Hampshire

CAPÍTULO 1

Introducción

1.1 INTRODUCCIÓN

En años recientes, el desarrollo de la estadística se ha hecho sentir en casi todas las fases de la actividad humana. La estadística ya no consiste meramente en la recopilación de datos y su presentación en gráficas y tablas; ahora se considera que abarca la ciencia de basar las inferencias sobre datos observados y la totalidad del problema de tomar decisiones en presencia de la incertidumbre. Esto cubre un terreno considerable puesto que nos encontramos con incertidumbres cuando lanzamos una moneda, cuando un dietista experimenta con aditivos para los alimentos, cuando un actuario determina las primas para el seguro de vida, cuando un ingeniero de control de calidad acepta o rechaza productos manufacturados, cuando un profesor compara las habilidades de los estudiantes, cuando un economista pronostica tendencias, cuando un periódico predice una elección, y así sucesivamente.

Sería presuntuoso decir que la estadística, en su estado actual de desarrollo, puede manejar todas las situaciones que implican incertidumbres, pero constantemente se desarrollan nuevas técnicas y la estadística moderna puede, al menos, proporcionar el marco de referencia para examinar estas situaciones en forma lógica y sistemática. En otras palabras, la estadística proporciona los modelos necesarios para estudiar las situaciones que implican incertidumbres, en la misma forma que el cálculo provee los modelos para describir, digamos, los conceptos de la física newtoniana.

Los orígenes de las matemáticas de la estadística se pueden encontrar en los estudios sobre probabilidad de mediados del siglo XIX, motivados por el interés en los juegos de azar. La teoría así desarrollada para "cara o cruz" o "rojo o negro" pronto encontró aplicaciones en situaciones donde los resultados eran "niño o niña", "vida o muerte" o "aprobar o reprobar", y los estudiosos empezaron a aplicar la teoría de la probabilidad a los problemas actuariales y a algunos aspectos de las ciencias sociales. Más tarde, L. Boltzmann, J. Gibbs y J. Maxwell introdujeron la probabilidad y la estadística a la física, y en este siglo se han encontrado aplicaciones en todas las fases del quehacer humano que en alguna forma implican un elemento de incertidumbre o riesgo. Los nombres relacionados de manera más prominente con el desarrollo de la estadística matemática en la primera mitad de este siglo son los de R. A. Fisher, J. Neyman, E. S. Pearson y A. Wald. Más recientemente, el trabajo de R. Schlaifer, L. J. Savage y otros más han dado ímpetu a las teorías estadísticas basadas, esencialmente, en métodos que se remontan al clérigo inglés Thomas Bayes del siglo XIX.

El enfoque a la estadística que se presenta en este libro es esencialmente el enfoque clásico, con métodos de inferencia ampliamente basados en el trabajo de J. Neyman y E. S. Pearson. Sin embargo, en el capítulo 9 se introduce el enfoque más general de la teoría de decisiones, y en el capítulo 10 se presentan algunos métodos Bayesianos. Este material se puede omitir sin que resulte una pérdida de continuidad.

1.2 MÉTODOS COMBINATORIOS

En muchos problemas de estadística debemos enumerar todas las alternativas posibles en una situación dada o al menos determinar cuántas posibilidades diferentes existen. En relación con esto último, a menudo usamos el siguiente teorema, algunas veces conocido como el **principio básico de conteo,** la **regla de conteo para eventos compuestos**, o la **regla de multiplicación de opciones.**

> **TEOREMA 1.1** Si una operación consta de dos pasos, de los cuales el primero se puede llevar a cabo en n_1 maneras y para cada una de éstas el segundo se puede hacer en n_2 maneras, entonces la operación completa se puede efectuar en $n_1 \cdot n_2$ maneras.

Aquí "operación" representa cualquier clase de procedimiento, proceso o método de selección.

Para justificar este teorema, definamos al par ordenado (x_i, y_j) como el resultado que surge cuando el primer paso resulta en la posibilidad x_i y el segundo paso en la posibilidad y_j. Entonces, el conjunto de todos los resultados posibles está formado por los siguientes $n_1 \cdot n_2$ pares:

$$\begin{gathered}
(x_1, y_1), (x_1, y_2), \ldots, (x_1, y_{n_2}) \\
(x_2, y_1), (x_2, y_2), \ldots, (x_2, y_{n_2}) \\
\cdots \\
\cdots \\
\cdots \\
(x_{n_1}, y_1), (x_{n_1}, y_2), \ldots, (x_{n_1}, y_{n_2})
\end{gathered}$$

EJEMPLO 1.1

Supongamos que alguien quiere ir de vacaciones en autobús, en tren o en avión por una semana a alguno de los cinco estados centrales del Noreste. Encuentre el número de maneras diferentes posibles de hacerlo.

Solución

El estado en particular se puede escoger de $n_1 = 5$ maneras y los medios de transporte se pueden escoger de $n_2 = 3$ maneras. Por consiguiente, el viaje se puede efectuar de $5 \cdot 3 = 15$ posibles maneras. Si se desea una lista de todas las posibilidades, un **diagrama de árbol** como el de la figura 1.1 proporciona un enfoque sistemático. Este diagrama muestra que hay $n_1 = 5$ ramas (posibilidades) para el número de estados, y para cada una de estas ramas hay $n_2 = 3$ ramas (posibilidades) para los diferentes medios de transporte. Es evidente que las 15 posibles maneras de tomar las vacaciones están representadas por los 15 trayectos distintos a lo largo de las ramas del árbol. ▲

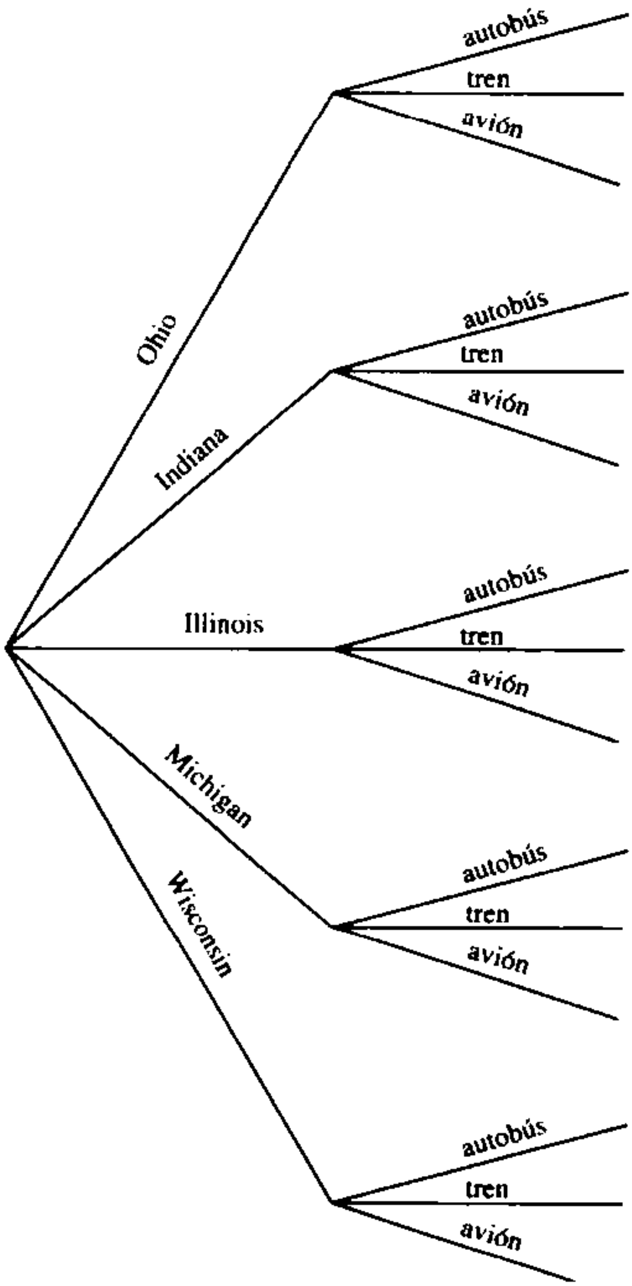

Figura 1.1 Diagrama de árbol.

EJEMPLO 1.2

¿Cuántos resultados posibles existen cuando tiramos un par de dados, uno rojo y uno verde?

Solución

El dado rojo puede caer en una de seis maneras, y para cada una de éstas el dado verde también puede caer de seis maneras. Por consiguiente el par de dados puede caer de $6 \cdot 6 = 36$ maneras. ▲

El teorema 1.1 se puede ampliar para abarcar situaciones donde una operación consta de dos o más pasos. En este caso

> **TEOREMA 1.2** Si una operación consta de k pasos, de los cuales el primero se puede llevar a cabo de n_1 maneras, para cada una de éstas el segundo paso se puede efectuar de n_2 maneras, para cada uno de los primeros dos el tercer paso se puede hacer en n_3 maneras, y así sucesivamente, entonces la operación completa se puede realizar en $n_1 \cdot n_2 \cdot \ldots \cdot n_k$ maneras.

EJEMPLO 1.3

Un inspector de control de calidad desea seleccionar una parte para la inspección de cada uno de cuatro recipientes diferentes que contienen 4, 3, 5 y 4 partes, respectivamente. ¿De cuántas maneras diferentes se pueden escoger las cuatro partes?

Solución

El número total de maneras es $4 \cdot 3 \cdot 5 \cdot 4 = 240$. ▲

EJEMPLO 1.4

¿De cuántas maneras diferentes se puede contestar todas las preguntas de una prueba de falso o verdadero que consta de 20 preguntas?

Solución

En total hay

$$2 \cdot 2 \cdot 2 \cdot 2 \cdot \ldots \cdot 2 \cdot 2 = 2^{20} = 1{,}048{,}576$$

maneras diferentes como se pueden responder todas las preguntas; sólo una de éstas corresponde al caso donde todas las respuestas son correctas y sólo una corresponde al caso donde todas las respuestas son incorrectas. ▲

Frecuentemente, estamos interesados en situaciones donde los resultados son las maneras diferentes en las que un grupo de objetos se pueden ordenar o arreglar. Por ejemplo, quizá deseamos saber de cuántas maneras diferentes los 24 miembros de un club pueden elegir a un presidente, un vicepresidente, un tesorero y un secretario, o podríamos desear saber de cuántas maneras diferentes seis personas se pueden sentar a la mesa. Los diferentes arreglos como éstos se conocen como **permutaciones**.

EJEMPLO 1.5

¿Cuántas permutaciones hay de las letras a, b y c?

Solución

Los arreglos posibles son abc, acb, bac, bca, cab y cba, así que el número de permutaciones diferentes es seis. Mediante el teorema 1.2 podríamos haber llegado a esta respuesta sin enumerar realmente las diferentes permutaciones. Puesto que hay tres opciones para seleccionar una letra para la primera posición, después dos para la segunda posición, dejando sólo una letra para la tercera posición, el número total de permutaciones es $3 \cdot 2 \cdot 1 = 6$. ▲

Al generalizar el argumento que se utilizó en el ejemplo anterior, encontramos que n objetos diferentes se pueden arreglar de $n(n-1)(n-2)\cdot \ldots \cdot 3 \cdot 2 \cdot 1$ maneras diferentes. Para simplificar nuestra notación, representamos este producto con el símbolo $n!$, el cual se lee "factorial n". Por lo tanto, $1! = 1$, $2! = 2 \cdot 1 = 2$, $3! = 3 \cdot 2 \cdot 1 = 6$, $4! = 4 \cdot 3 \cdot 2 \cdot 1 = 24$, $5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$ y así sucesivamente. También, por definición $0! = 1$.

TEOREMA 1.3 El número de permutaciones de n objetos diferentes es $n!$.

EJEMPLO 1.6

¿De cuántas maneras diferentes se pueden presentar al público los cinco jugadores titulares de un equipo de baloncesto?

Solución

Hay $5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$ maneras en las que se pueden presentar. ▲

EJEMPLO 1.7

El número de permutaciones de las cuatro letras a, b, c y d es 24, pero ¿cuál es el número de permutaciones si sólo tomamos dos de las cuatro letras o, como usualmente se expresa, si tomamos las cuatro letras dos a la vez?

Solución

Tenemos dos posiciones que llenar, con cuatro opciones para la primera y después tres opciones para la segunda. Por consiguiente, mediante el teorema 1.1, el número de permutaciones es $4 \cdot 3 = 12$. ▲

Al generalizar el argumento que usamos en el ejemplo anterior, encontramos que n objetos diferentes tomados r a la vez, para $r > 0$, se pueden arreglar de

$n(n-1)\cdot \ldots \cdot(n-r+1)$ maneras. Se representa este producto mediante ${}_nP_r$, y hacemos ${}_nP_0 = 1$ por definición. Por consiguiente, podemos escribir lo siguiente:

> **TEOREMA 1.4** El número de permutaciones de n objetos diferentes tomados r a la vez es
>
> $$ {}_nP_r = \frac{n!}{(n-r)!} $$
>
> para $r = 0, 1, 2, \ldots, n$.

Demostración. La fórmula ${}_nP_r = n(n-1)\cdot \ldots \cdot(n-r+1)$ no se puede usar para $r = 0$, pero tenemos

$$ {}_nP_0 = \frac{n!}{(n-0)!} = 1 $$

Para $r = 1, 2, \ldots, n$, tenemos

$$ \begin{aligned} {}_nP_r &= n(n-1)(n-2)\cdot \ldots \cdot(n-r+1) \\ &= \frac{n(n-1)(n-2)\cdot \ldots \cdot(n-r+1)(n-r)!}{(n-r)!} \\ &= \frac{n!}{(n-r)!} \qquad \blacktriangledown \end{aligned} $$

En los problemas concernientes a permutaciones, suele ser más fácil proceder con el uso del teorema 1.2 como en el ejemplo 1.7, pero la fórmula factorial del teorema 1.4 es más fácil de recordar. Muchos paquetes de software de estadística proporcionan valores para ${}_nP_r$ y otras cantidades combinatorias mediante sencillas instrucciones. De hecho, estas cantidades también están preprogramadas en muchas calculadoras manuales de estadística (o científicas).

EJEMPLO 1.8

De entre los 24 miembros de un club se sacan cuatro nombres para los puestos de presidente, vicepresidente, tesorero y secretario. ¿De cuántas maneras diferentes se puede hacer esto?

Solución

El número de permutaciones de 24 objetos diferentes tomados 4 a la vez es

$$ {}_{24}P_4 = \frac{24!}{20!} = 24\cdot 23\cdot 22\cdot 21 = 255{,}024 \qquad \blacktriangle $$

EJEMPLO 1.9

¿De cuántas maneras puede una sección local de la Sociedad Americana de Química programar a tres oradores para tres reuniones diferentes, si todos ellos están disponibles en cualquiera de cinco fechas posibles?

Solución

Puesto que debemos escoger tres de cinco fechas e importa el orden en que se escogen (asignadas a los tres oradores), obtenemos

$$_5P_3 = \frac{5!}{2!} = \frac{120}{2} = 60$$

También podemos argumentar que el primer orador se puede programar en cualquiera de cinco maneras, el segundo orador de cuatro maneras y el tercer orador de tres maneras, de modo que la respuesta es $5 \cdot 4 \cdot 3 = 60$. ▲

Las permutaciones que ocurren cuando los objetos se arreglan en un círculo se llaman **permutaciones circulares**. Dos permutaciones circulares no se consideran diferentes (y se cuentan sólo una vez) si los objetos correspondientes en los dos arreglos tienen los mismos objetos a su izquierda y a su derecha. Por ejemplo, si cuatro personas están jugando bridge, no obtenemos una permutación diferente si todos se cambian a la silla que está a su derecha.

EJEMPLO 1.10

¿Cuántas permutaciones circulares hay de cuatro personas que juegan bridge?

Solución

Si consideramos arbitrariamente la posición de uno de los cuatro jugadores como fija, podemos sentar (arreglar) a los otros tres jugadores en $3! = 6$ maneras diferentes. En otras palabras, hay seis permutaciones circulares diferentes. ▲

Al generalizar el argumento que se utilizó en el ejemplo anterior, obtenemos el siguiente teorema.

> **TEOREMA 1.5** El número de permutaciones de n objetos diferentes arreglados en un círculo es $(n - 1)!$.

Hasta ahora hemos supuesto que los n objetos, de los que seleccionamos r objetos y formamos permutaciones, son todos diferentes. Así, por ejemplo, no se pueden usar las diversas fórmulas para determinar el número de maneras en las que podemos arreglar las letras en la palabra "*book*" (libro) o de cuántas maneras se pueden arreglar tres copias de una novela y una copia de otras cuatro en un entrepaño.

EJEMPLO 1.11

¿Cuántas permutaciones diferentes hay de las letras de la palabra "*book*"?

Solución

Si por el momento distinguimos entre las dos o etiquetándolas o_1 y o_2, entonces hay $4! = 24$ permutaciones diferentes de los símbolos b, o_1, o_2 y k. Sin embargo, si quitamos los subíndices, entonces bo_1ko_2 y bo_2ko_1, por ejemplo, ambos dan como resultado *boko*, y puesto que cada par de permutaciones con subíndice resulta en sólo un arreglo sin subíndices, el número total de arreglos de letras en la palabra "*book*" es $\frac{24}{2} = 12$. ▲

EJEMPLO 1.12

¿De cuántas maneras diferentes se pueden arreglar, en un entrepaño, tres copias de una novela y una copia de cada una de otras cuatro novelas?

Solución

Si designamos las tres copias de la primera novela con a_1, a_2 y a_3, y las otras cuatro novelas con b, c, d y e, encontramos que *con subíndices* hay 7! diferentes permutaciones de a_1, a_2, a_3, b, c, d y e. Sin embargo, puesto que hay 3! permutaciones de a_1, a_2 y a_3 que llevan a la misma permutación de a, a, a, b, c, d y e, encontramos que sólo hay $\frac{7!}{3!} = 7 \cdot 6 \cdot 5 \cdot 4 = 840$ maneras en las que siete libros se pueden arreglar en un entrepaño. ▲

Al generalizar el argumento utilizado en estos dos ejemplos, obtenemos el siguiente teorema.

TEOREMA 1.6 El número de permutaciones de n objetos de los cuales n_1 son de una clase, n_2 son de una segunda clase,..., n_k son de la $k$ésima clase y $n_1 + n_2 + \cdots + n_k = n$ es $n_1 + n_2$

$$\frac{n!}{n_1! \cdot n_2! \cdot \ldots \cdot n_k!}$$

EJEMPLO 1.13

¿De cuántas maneras se pueden colgar, una junto a la otra, dos pinturas de Monet, tres pinturas de Renoir y dos pinturas de Degas en la pared de un museo sin hacer distinción entre las pinturas de los mismos artistas?

Solución

Al sustituir $n = 7$, $n_1 = 2$, $n_2 = 3$ y $n_3 = 2$ en la fórmula del teorema 1.6, obtenemos

$$\frac{7!}{2! \cdot 3! \cdot 2!} = 210 \qquad \blacktriangle$$

Hay muchos problemas en los que nos interesa determinar el número de maneras en las cuales r objetos se pueden seleccionar de entre n objetos diferentes *sin importar el orden en el cual son seleccionados*. Tales selecciones (arreglos) se conocen como **combinaciones**.

EJEMPLO 1.14

¿De cuántas maneras diferentes puede una persona, que reúne datos para una organización de investigación de mercados, seleccionar tres de 20 familias que viven en un complejo departamental dado?

Solución

Si nos interesa el orden en el cual se selecciona a las familias, la respuesta es

$${}_{20}P_3 = 20 \cdot 19 \cdot 18 = 6{,}840$$

pero entonces cada conjunto de tres familias se contaría $3! = 6$ veces. Si no nos interesa el orden en que se seleccionan las familias, sólo hay $\frac{6{,}840}{6} = 1{,}140$ maneras en que la persona que reúne los datos puede hacer su trabajo. $\blacktriangle$

Realmente, "combinación" significa lo mismo que "subconjunto", y cuando pedimos el número de combinaciones de r objetos seleccionados de un conjunto de n objetos diferentes, simplemente pedimos el número total de subconjuntos de r objetos que se pueden seleccionar de un conjunto de n objetos diferentes. En general, hay $r!$ permutaciones de los objetos en un subconjunto de r objetos, así que las ${}_nP_r$ permutaciones de r objetos seleccionados de un conjunto de n objetos diferentes contienen cada subconjunto $r!$ veces. Al dividir ${}_nP_r$ por $r!$ y representar el resultado por medio del símbolo $\binom{n}{r}$, tenemos entonces

> **TEOREMA 1.7** El número de combinaciones de n objetos diferentes tomados r a la vez es
>
> $$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$
>
> para $r = 0, 1, 2, \ldots, n$.

EJEMPLO 1.15

¿En cuántas formas diferentes pueden seis lanzamientos de una moneda, producir dos caras y cuatro cruces?

Solución

Esta pregunta es lo mismo que preguntar de cuántas maneras podemos seleccionar los dos lanzamientos, en los cuales ocurrirán caras. Por consiguiente, al aplicar el teorema 1.7, encontramos que la respuesta es

$$\binom{6}{2} = \frac{6!}{2! \cdot 4!} = 15$$

También se puede obtener este resultado con el proceso, bastante tedioso, de enumerar las diversas posibilidades, HHTTTT, TTHTHT, HTHTTT, ..., donde H representa cara y T representa cruz. ▲

EJEMPLO 1.16

¿Cuántos comités diferentes, de dos químicos y un físico, se pueden formar con los cuatro químicos y los tres físicos del profesorado de una pequeña universidad?

Solución

Puesto que dos de los cuatro químicos se pueden seleccionar de $\binom{4}{2} = \frac{4!}{2! \cdot 2!} = 6$ maneras y uno de los tres físicos se puede seleccionar de $\binom{3}{1} = \frac{3!}{1! \cdot 2!} = 3$ maneras, el teorema 1.1 muestra que el número de comités es $6 \cdot 3 = 18$. ▲

Una combinación de r objetos seleccionados de un conjunto de n objetos diferentes se puede considerar una **partición** de los n objetos en dos subconjuntos que contienen, respectivamente, los r objetos que se seleccionan y los $n - r$ objetos restantes. A menudo, nos centramos en el problema más general de dividir un conjunto de n objetos distintos en k subconjuntos, lo cual requiere que cada uno de los n objetos debe pertenecer a uno y sólo a uno de los subconjuntos.† No importa el orden de los objetos dentro de un subconjunto.

EJEMPLO 1.17

¿De cuántas maneras se puede dividir un conjunto de cuatro objetos en tres subconjuntos que contengan, respectivamente, dos, uno y uno de los objetos?

Solución

A representar los cuatro objetos por a, b, c y d, encontramos, por enumeración, que existen las siguientes 12 posibilidades:

---

† Simbólicamente, los subconjuntos $A_1, A_2, \ldots, A_k$ constituyen una partición del conjunto A si $A_1 \cup A_2 \cup \cdots \cup A_k = A$ y $A_i \cap A_j = \emptyset$ para toda $i \neq j$.

$$\begin{array}{cccc} ab|c|d & ab|d|c & ac|b|d & ac|d|b \\ ad|b|c & ad|c|b & bc|a|d & bc|d|a \\ bd|a|c & bd|c|a & cd|a|b & cd|b|a \end{array}$$

El número de particiones para este ejemplo se representa con el símbolo

$$\binom{4}{2,1,1} = 12$$

donde el número de la parte superior representa el número total de objetos y los números de la parte inferior denotan el número de objetos que entran en cada subconjunto. ▲

Si no hubiéramos querido enumerar todas las posibilidades en el ejemplo precedente, podríamos haber argumentado que los dos objetos que entran en el primer subconjunto se pueden escoger de $\binom{4}{2} = 6$ maneras, el objeto que entra en el segundo subconjunto puede entonces elegirse de $\binom{2}{1} = 2$ maneras y el objeto que entra en el tercer subconjunto puede entonces elegirse de $\binom{1}{1} = 1$ maneras. Así, por medio del teorema 1.2, hay $6 \cdot 2 \cdot 1 = 12$ particiones. Al generalizar este argumento tenemos el siguiente teorema.

TEOREMA 1.8 El número de maneras en que un conjunto de n objetos diferentes se pueden dividir o partir en k subconjuntos de n_1 objetos en el primer subconjunto, n_2 objetos en el segundo subconjunto,..., y n_k objetos en el $k$ésimo subconjunto es

$$\binom{n}{n_1, n_2, \ldots, n_k} = \frac{n!}{n_1! \cdot n_2! \cdot \ldots \cdot n_k!}$$

Demostración. Puesto que los n_1 objetos que entran en el primer subconjunto se pueden escoger de $\binom{n}{n_1}$ maneras, los n_2 objetos que entran en el segundo subconjunto pueden entonces escogerse de $\binom{n-n_1}{n_2}$ maneras, los n objetos que entran en el tercer subconjunto pueden entonces escogerse de $\binom{n-n_1-n_2}{n_3}$ maneras y así sucesivamente, resulta por el teorema 1.2 que el número total de particiones es

$$\begin{aligned}\binom{n}{n_1, n_2, \ldots, n_k} &= \binom{n}{n_1}\cdot\binom{n-n_1}{n_2}\cdot\ldots\cdot\binom{n-n_1-n_2-\cdots-n_{k-1}}{n_k}\\ &= \frac{n!}{n_1!\cdot(n-n_1)!}\cdot\frac{(n-n_1)!}{n_2!\cdot(n-n_1-n_2)!}\\ &\qquad\cdot\ldots\cdot\frac{(n-n_1-n_2-\cdots-n_{k-1})!}{n_k!\cdot 0!}\\ &= \frac{n!}{n_1!\cdot n_2!\cdot\ldots\cdot n_k!} \qquad \blacktriangledown\end{aligned}$$

EJEMPLO 1.18

¿De cuántas maneras puede asignarse a siete hombres de negocios, que asisten a una convención, una habitación triple de hotel y dos dobles?

Solución

Sustituyendo $n = 7$, $n_1 = 3$, $n_2 = 2$ y $n_3 = 2$ en la fórmula del teorema 1.8, obtenemos

$$\binom{7}{3, 2, 2} = \frac{7!}{3!\cdot 2!\cdot 2!} = 210 \qquad \blacktriangle$$

1.3 COEFICIENTES BINOMIALES

Si n es un entero positivo y multiplicamos $(x + y)^n$ término por término, cada término será el producto de las x y las y, donde una x o una y proviene de cada uno de los n factores $x + y$. Por ejemplo, la expansión

$$\begin{aligned}(x+y)^3 &= (x+y)(x+y)(x+y)\\ &= x\cdot x\cdot x + x\cdot x\cdot y + x\cdot y\cdot x + x\cdot y\cdot y\\ &\quad + y\cdot x\cdot x + y\cdot x\cdot y + y\cdot y\cdot x + y\cdot y\cdot y\\ &= x^3 + 3x^2y + 3xy^2 + y^3\end{aligned}$$

produce términos de la forma x^3, x^2y, xy^2 y y^3. Sus coeficientes son 1, 3, 3 y 1, y el coeficiente de xy^2, por ejemplo, es $\binom{3}{2} = 3$, el número de maneras en que podemos escoger los dos factores que proveen las y. En forma similar, el coeficiente de x^2y es $\binom{3}{1} = 3$, el número de maneras en que podemos escoger el único factor que provee la y, y los coeficientes de x^3 y y^3 son $\binom{3}{0} = 1$ y $\binom{3}{3} = 1$.

Más generalmente, si n es un entero positivo y multiplicamos $(x+y)^n$ término por término, el coeficiente de $x^{n-r}y^r$ es $\binom{n}{r}$, el número de maneras en que podemos escoger r factores que provean las y. En consecuencia, nos referimos a $\binom{n}{r}$ como un **coeficiente binomial**. Ahora podemos enunciar el siguiente teorema.

TEOREMA 1.9

$$(x+y)^n = \sum_{r=0}^{n} \binom{n}{r} x^{n-r}y^r \qquad \text{para cualquier entero positivo } n$$

(Para los lectores que no estén familiarizados con la notación $\sum$ en el apéndice A se da una breve explicación.)

A menudo se puede simplificar el cálculo de los coeficientes binomiales al utilizar los tres teoremas siguientes.

TEOREMA 1.10 Para cualesquier enteros positivos n y $r = 0, 1, 2, \ldots, n$,

$$\binom{n}{r} = \binom{n}{n-r}$$

Demostración. Podríamos argumentar que cuando seleccionamos un subconjunto de r objetos de un conjunto de n objetos diferentes, dejamos un subconjunto de $n-r$ objetos; de ahí que, hay tantas maneras de seleccionar r objetos como maneras de dejar (o seleccionar) $n-r$ objetos. Para demostrar el teorema en forma algebraica, escribimos

$$\binom{n}{n-r} = \frac{n!}{(n-r)![n-(n-r)]!} = \frac{n!}{(n-r)!r!}$$

$$= \frac{n!}{r!(n-r)!} = \binom{n}{r} \qquad \blacktriangledown$$

El teorema 1.10 implica que si calculamos los coeficientes binomiales para $r = 0, 1, \ldots, \frac{n}{2}$ cuando n es par y para $r = 0, 1, \ldots, \frac{n-1}{2}$, cuando n es impar, los coeficientes binomiales restantes se pueden obtener al utilizar el teorema.

EJEMPLO 1.19

Dado $\binom{4}{0} = 1$, $\binom{4}{1} = 4$, y $\binom{4}{2} = 6$, encuentre $\binom{4}{3}$ y $\binom{4}{4}$

Solución

$$\binom{4}{3} = \binom{4}{4-3} = \binom{4}{1} = 4 \quad \text{y} \quad \binom{4}{4} = \binom{4}{4-4} = \binom{4}{0} = 1 \qquad \blacktriangle$$

EJEMPLO 1.20

Dado $\binom{5}{0} = 1$, $\binom{5}{1} = 5$, y $\binom{5}{2} = 10$, encuentre $\binom{5}{3}$, $\binom{5}{4}$, y $\binom{5}{5}$.

Solución

$$\binom{5}{3} = \binom{5}{5-3} = \binom{5}{2} = 10, \binom{5}{4} = \binom{5}{5-4} = \binom{5}{1} = 5, \text{ y}$$
$$\binom{5}{5} = \binom{5}{5-5} = \binom{5}{0} = 1. \qquad \blacktriangle$$

Es precisamente en esta forma como el teorema 1.10 podrá usarse en relación con la tabla VII.†

EJEMPLO 1.21

Encuentre $\binom{20}{12}$ y $\binom{17}{10}$.

Solución

Puesto que $\binom{20}{12}$ no se da en la tabla VII, utilizamos el hecho de que $\binom{20}{12} = \binom{20}{8}$, buscamos $\binom{20}{8}$, y obtenemos $\binom{20}{12} = 125{,}970$. Asimismo, para encontrar $\binom{17}{10}$, utilizamos el hecho de que $\binom{17}{10} = \binom{17}{7}$, buscamos $\binom{17}{7}$, y obtenemos $\binom{17}{10} = 19{,}448$. $\blacktriangle$

TEOREMA 1.11 Para cualquier entero positivo n y $r = 1, 2, \ldots, n-1$,

$$\binom{n}{r} = \binom{n-1}{r} + \binom{n-1}{r-1}$$

† Los números romanos se refieren a las tablas estadísticas al final del libro.

Demostración. Al sustituir $x = 1$ en $(x + y)^n$, escribamos

$$(1 + y)^n = (1 + y)(1 + y)^{n-1} = (1 + y)^{n-1} + y(1 + y)^{n-1}$$

e igualemos el coeficiente de y^r en $(1 + y)^n$ con aquellos en $(1 + y)^{n-1} + y(1 + y)^{n-1}$. Puesto que el coeficiente de y^r en $(1 + y)^n$ es $\binom{n}{r}$ y el coeficiente de y^r en $(1 + y)^{n-1} + y(1 + y)^{n-1}$ es la suma del coeficiente de y^r en $(1 + y)^{n-1}$, esto es, $\binom{n-1}{r}$, y el coeficiente de y^{r-1} en $(1 + y)^{n-1}$, esto es, $\binom{n-1}{r-1}$, obtenemos

$$\binom{n}{r} = \binom{n-1}{r} + \binom{n-1}{r-1}$$

lo cual completa la demostración. ▼

Alternativamente, tome cualquiera de los n objetos. Si no incluirá entre los r objetos, hay $\binom{n-1}{r}$ maneras de seleccionar r objetos; si va a incluirse, hay $\binom{n-1}{r-1}$ maneras de seleccionar los otros r 2 1 objetos. Por consiguiente, hay $\binom{n-1}{r} + \binom{n-1}{r-1}$ maneras de seleccionar los r objetos, esto es,

$$\binom{n}{r} = \binom{n-1}{r} + \binom{n-1}{r-1}$$

El teorema 1.11 también se puede demostrar al expresar los coeficientes binomiales, en ambos lados de la ecuación, en términos de factoriales y entonces proceder de manera algebraica, pero dejaremos esto al lector en el ejercicio 1.12. En el ejercicio 1.11 se da una aplicación importante del teorema 1.11, donde proporciona la clave para la construcción de lo que se conoce como el **triángulo de Pascal**.

Para enunciar el tercer teorema sobre los coeficientes binomiales, hagamos la siguiente definición: $\binom{n}{r} = 0$ siempre que n sea un entero positivo y r sea un entero positivo mayor que n. (Evidentemente, no hay forma en que podamos seleccionar un subconjunto que contenga más elementos que todo el conjunto mismo.)

TEOREMA 1.12

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

Demostración. Usando la misma técnica que en la demostración del teorema 1.11, demostremos este teorema al igualar los coeficientes de y^k en las expresiones de ambos lados de la ecuación

$$(1 + y)^{m+n} = (1 + y)^m(1 + y)^n$$

El coeficiente de y^k en $(1 + y)^{m+n}$ es $\binom{m+n}{k}$, y el coeficiente de y^k en

$$(1 + y)^m(1 + y)^n = \left[\binom{m}{0} + \binom{m}{1}y + \cdots + \binom{m}{m}y^m\right]$$
$$\times \left[\binom{n}{0} + \binom{n}{1}y + \cdots + \binom{n}{n}y^n\right]$$

es la suma de los productos que obtenemos al multiplicar el término constante del primer factor por el coeficiente de y^k en el segundo factor, el coeficiente de y en el primer factor por el coeficiente de y^{k-1} en el segundo factor, ..., y el coeficiente de y^k en el primer factor por el término constante del segundo factor. Así, el coeficiente de y^k en $(1 + y)^m(1 + y)^n$ es

$$\binom{m}{0}\binom{n}{k} + \binom{m}{1}\binom{n}{k-1} + \binom{m}{2}\binom{n}{k-2} + \cdots + \binom{m}{k}\binom{n}{0}$$
$$= \sum_{r=0}^{k}\binom{m}{r}\binom{n}{k-r}$$

y esto completa la demostración. ▼

EJEMPLO 1.22

Verifique el teorema 1.12 numéricamente para $m = 2$, $n = 3$ y $k = 4$.

Solución

Al sustituir estos valores, obtenemos

$$\binom{2}{0}\binom{3}{4} + \binom{2}{1}\binom{3}{3} + \binom{2}{2}\binom{3}{2} + \binom{2}{3}\binom{3}{1} + \binom{2}{4}\binom{3}{0} = \binom{5}{4}$$

y puesto que $\binom{3}{4}$, $\binom{2}{3}$ y $\binom{2}{4}$ son igual a 0 de acuerdo a la definición en la página 15, la ecuación se reduce a

$$\binom{2}{1}\binom{3}{3} + \binom{2}{2}\binom{3}{2} = \binom{5}{4}$$

lo cual se comprueba, puesto que $2 \cdot 1 + 1 \cdot 3 = 5$. ▲

Al utilizar el teorema 1.8, podemos ampliar nuestra exposición a **coeficientes multinomiales**, esto es, a los coeficientes que resultan de la expansión de $(x_1 + x_2$

$+ \cdots + x_k)^n$. El coeficiente multinomial del término $x_1^{r_1} \cdot x_2^{r_2} \cdot \ldots \cdot x_k^{r_k}$ en la expansión de $(x_1 + x_2 + \cdots + x_k)^n$ es

$$\binom{n}{r_1, r_2, \ldots, r_k} = \frac{n!}{r_1! \cdot r_2! \cdot \ldots \cdot r_k!}$$

EJEMPLO 1.23

¿Cuál es el coeficiente de $x_1^3 x_2 x_3^2$ en la expansión de $(x_1 + x_2 + x_3)^6$?

Solución

Si sustituimos $n = 6$, $r_1 = 3$, $r_2 = 1$ y $r_3 = 2$ en la fórmula anterior, obtenemos

$$\frac{6!}{3! \cdot 1! \cdot 2!} = 60 \qquad \blacktriangle$$

EJERCICIOS

1.1 Una operación consta de dos pasos, de los cuales el primero se puede hacer de n_1 maneras. Si el primer paso se hace de la manera $i$ésima, el segundo paso se puede hacer de n_{2i} maneras.†

(a) Use un diagrama de árbol y encuentre una fórmula para el número total de maneras en que se puede efectuar la operación total.

(b) Un estudiante puede prepararse durante 0, 1, 2 o 3 horas para un examen de historia en un día dado. Use la fórmula obtenida en la parte (a) para verificar que hay 13 maneras en las que el estudiante puede prepararse durante 4 horas cuando mucho para la prueba en dos días consecutivos.

1.2 Con respecto al ejercicio 1.1 verifique que si n_{2i} es igual a la constante n_2, la fórmula obtenida en la parte (a) se reduce a aquella del teorema 1.1.

1.3 Con respecto al ejercicio 1.1, supongamos que hay un tercer paso, y si el primer paso se realizó de la $i$ésima manera y el segundo paso de la $j$ésima manera, el tercer paso se puede hacer de n_{3ij} maneras.

(a) Use un diagrama de árbol para verificar que toda la operación se puede hacer de

$$\sum_{i=1}^{n_1} \sum_{j=1}^{n_{2i}} n_{3ij}$$

maneras diferentes.

(b) Con respecto al inciso (b) del ejercicio 1.1, use la fórmula del inciso (a) para verificar que hay 32 maneras en las que el estudiante puede prepararse durante 4 horas, cuando mucho, para la prueba en tres días consecutivos.

1.4 Demuestre que si n_{2i} es igual a la constante n_2 y n_{3ij} es igual a la constante n_3, la fórmula del inciso (a) del ejercicio 1.3 se reduce a la del teorema 1.2.

† El uso de subíndices dobles se explica en el apéndice A.

1.5 En una serie final del campeonato entre dos equipos de baloncesto, el ganador es el primer equipo que gane m juegos.

(a) Contando separadamente el número de series finales que requieren m, $m + 1, \ldots$, y $2m - 1$ juegos, muestre que el número total de resultados diferentes (secuencias de juegos ganados y juegos perdidos de uno de los equipos) es

$$2\left[\binom{m-1}{m-1} + \binom{m}{m-1} + \cdots + \binom{2m-2}{m-1}\right]$$

(b) ¿Cuántos resultados diferentes hay para una final de "2 de 3", una final de "3 de 5" y una final de "4 de 7"?

1.6 Cuando n es grande, se puede aproximar $n!$ por medio de la expresión

$$\sqrt{2\pi n}\left(\frac{n}{e}\right)^n$$

llamada la **fórmula de Stirling**, donde e es la base de los logaritmos naturales. (Se puede encontrar una derivación de esta fórmula en el libro de W. Feller citado entre las referencias al final de este capítulo.)

(a) Utilice la fórmula de Stirling y obtenga las aproximaciones para $10!$ y $12!$, también encuentre los porcentajes de error de estas aproximaciones al compararlas con los valores exactos, dados en la tabla VII.

(b) Use la fórmula de Stirling y obtenga una aproximación para el número de manos de bridge de 13 cartas que se pueden dar con una baraja ordinaria de 52 cartas de juego.

1.7 Use la fórmula de Stirling (véase el ejercicio 1.6) para aproximar $2n!$ y $n!$, muestre que

$$\frac{\binom{2n}{n}\sqrt{\pi n}}{2^{2n}} \approx 1$$

1.8 En algunos problemas de la **teoría de la ocupación** nos interesa el número de maneras en que ciertos objetos *distinguibles* se pueden distribuir entre personas, urnas, cajas o celdas. Encuentre una expresión para el número de formas en que se pueden distribuir r objetos *distinguibles* entre n celdas y úsela para encontrar el número de maneras en que se pueden distribuir tres libros diferentes entre 12 estudiantes, en una clase de literatura inglesa.

1.9 En algunos problemas de la teoría de la ocupación nos interesa de cuántas maneras se pueden distribuir ciertos objetos *indistinguibles* entre personas, urnas, cajas o celdas. Encuentre una expresión para el número de maneras en que se pueden distribuir r objetos *indistinguibles* entre n celdas y úsela para encontrar el número de maneras en que un panadero puede vender cinco hogazas (indistinguibles) de pan a tres clientes. (*Sugerencia*: Podríamos argumentar que L|LLL|L presenta el caso donde los tres clientes compran una hogaza, tres hogazas, y una hogaza, respectivamente, y que LLLL| |L representa el caso donde los tres clientes compran cuatro hogazas, ninguna hogaza y una hogaza. Así, debemos buscar el número de maneras en que podemos arreglar las cinco H y las dos barras verticales.)

1.10 En algunos problemas de la teoría de la ocupación nos interesa el número de maneras en que ciertos objetos *indistinguibles* se pueden distribuir entre individuos, urnas, cajas o celdas con al menos uno en cada celda. Encuentre una expresión para el número de maneras en que r objetos *indistinguibles* se pueden distribuir entre n celdas con al menos una en cada celda y vuelva a trabajar en la parte numérica del ejercicio 1.9, donde cada uno de los tres clientes obtiene al menos una hogaza de pan.

1.11 Cuando no hay tablas disponibles, a veces es conveniente determinar los coeficientes binomiales por medio del siguiente arreglo, llamado **triángulo de Pascal**:

$$\begin{array}{ccccccccccc}
 & & & & & 1 & & & & & \\
 & & & & 1 & & 1 & & & & \\
 & & & 1 & & 2 & & 1 & & & \\
 & & 1 & & 3 & & 3 & & 1 & & \\
 & 1 & & 4 & & 6 & & 4 & & 1 & \\
1 & & 5 & & 10 & & 10 & & 5 & & 1 \\
 & & & & \ldots & \ldots & \ldots & & & &
\end{array}$$

donde cada hilera empieza con un 1, termina con un 1, y cada uno de los demás elementos es la suma de los dos elementos más cercanos de la hilera que está inmediatamente arriba. En este triángulo, el elemento $r$ésimo en la $n$ésima hilera es el coeficiente binomial $\binom{n-1}{r-1}$. Construya las dos hileras siguientes (séptima y octava) del triángulo y escriba las expresiones binomiales de $(x+y)^6$ y $(x+y)^7$.

1.12 Demuestre el teorema 11.1 mediante la expresión de todos los coeficientes binomiales en términos de factoriales y después simplifique en forma algebraica.

1.13 Al expresar los coeficientes binomiales en términos de factoriales y simplificar en forma algebraica, demuestre que

(a) $\binom{n}{r} = \frac{n-r+1}{r}\cdot\binom{n}{r-1}$;

(b) $\binom{n}{r} = \frac{n}{n-r}\cdot\binom{n-1}{r}$;

(c) $n\binom{n-1}{r} = (\ \ + 1)\binom{n}{r+1}$.

1.14 Sustituya los valores apropiados para x y y en la fórmula del teorema 1.9, para demostrar que

(a) $\sum_{r=0}^{n}\binom{n}{r} = 2^n$;

(b) $\sum_{r=0}^{n}(-1)^r\binom{n}{r} = 0$;

(c) $\sum_{r=0}^{n}\binom{n}{r}\left(- 1\right)^{r} = a^{n}.$

1.15 Por medio de la aplicación repetida del teorema 1.11, demuestre que

$$\binom{n}{r} = \sum_{i=1}^{r+1}\binom{n-i}{r-i+1}$$

1.16 Use el teorema 1.12 para demostrar que

$$\sum_{r=0}^{n}\binom{n}{r}^{2} = \binom{2n}{n}$$

1.17 Demuestre que $\sum_{r=0}^{n} r\binom{n}{r} = n2^{n-1}$ haciendo $x = 1$ en el teorema 1.9, después diferenciando las expresiones en ambos lados con respecto a y, y finalmente sustituyendo $y = 1$.

1.18 Vuelva a trabajar en el ejercicio 1.17 usando el inciso (a) del ejercicio 1.14 y el inciso (c) del ejercicio 1.13.

1.19 Si n no es un entero positivo o cero, la expansión binomial de $(1 + y)^n$ produce, para $-1 < y < 1$, la serie infinita

$$1 + \binom{n}{1}y + \binom{n}{2}y^{2} + \binom{n}{3}y^{3} + \cdots + \binom{n}{r}y^{r} + \cdots$$

donde $\binom{n}{r} = \frac{n(n-1)\cdot\ldots\cdot(n-r+1)}{r!}$ para $r = 1, 2, 3, \ldots$. Use **esta definición generalizada de coeficientes binomiales** (que concuerda con la de la página 13 para valores enteros positivos de n) para evaluar

(a) $\binom{\frac{1}{2}}{4}$ y $\binom{-3}{3}$;

(b) $\sqrt{5}$ escribiendo $\sqrt{5} = 2(1 + \frac{1}{4})^{1/2}$ y usando los primeros cuatro términos de la expansión binomial de $(1 + \frac{1}{4})^{1/2}$.

1.20 Con respecto a la definición generalizada de los coeficientes binomiales en el ejercicio 1.19, demuestre que

(a) $\binom{-1}{r} = \left(-1\right)^{r}$;

(b) $\binom{-n}{r} = \left(-1\right)^{r}\binom{n+r-1}{r}$ para $n > 0$.

1.21 Encuentre el coeficiente de $x^2y^3z^3$ en la expansión de $(x + y + z)^8$.

1.22 Obtenga el coeficiente de $x^3y^2z^3w$ en la expansión de $(2x + 3y - 4z + w)^9$.

1.23 Demuestre que

$$\binom{n}{n_1, n_2, \ldots, n_k} = \binom{n-1}{n_1 - 1, n_2, \ldots, n_k} + \binom{n-1}{n_1, n_2 - 1, \ldots, n_k} + \cdots + \binom{n-1}{n_1, n_2, \ldots, n_k - 1}$$

expresando todos estos coeficientes multinomiales en términos de factoriales y simplificando en forma algebraica.

APLICACIONES

1.24 Hay cuatro rutas, A, B, C y D, entre la casa de una persona y el lugar donde trabaja, pero la ruta B es de un solo sentido, de modo que no puede tomarla cuando va a su trabajo, y la ruta C es de un solo sentido, de modo que no puede tomarla cuando va rumbo a casa.

(a) Trace un diagrama de árbol que muestre las diversas maneras en que la persona puede ir y venir del trabajo.

(b) Trace un diagrama de árbol que muestre las diversas maneras en que puede ir y venir del trabajo, sin tomar la misma ruta en ambos sentidos.

1.25 Una persona con \$2 en su bolsillo apuesta \$1, contra la misma cantidad, en un "volado" o lanzamiento de una moneda y continúa apostando \$1 en tanto tiene dinero. Trace un diagrama de árbol para mostrar las diversas situaciones que pueden suceder durante los primeros cuatro lanzamientos de la moneda. Después del cuarto lanzamiento, ¿en cuántos casos estará

(a) exactamente sin ganar ni perder;

(b) exactamente adelante por \$2?

1.26 Suponga que en la Serie Mundial de béisbol (en la cual el ganador es el primer equipo que gana cuatro juegos) el campeón de la Liga Nacional aventaja al campeón de la Liga Americana por tres juegos a dos. Construya un diagrama de árbol para mostrar de cuántas maneras pueden ganar o perder estos equipos el juego o los juegos restantes.

1.27 El entrenador de un campo de golf almacena dos juegos idénticos de palos de golf para mujer, reordenando al final de cada día (a fin de entregar temprano en la mañana siguiente) si y sólo si vendió ambos. Trace un diagrama de árbol para demostrar que si él empieza un lunes con dos juegos de palos, hay en total ocho diferentes maneras en que puede vender durante los primeros dos días de esa semana.

1.28 Por muchos siglos, contar el número de resultados en los juegos de azar ha sido un pasatiempo popular. Esto era de interés no sólo porque el juego (por dinero) estaba de por medio, sino también porque los resultados de los juegos de azar a menudo se interpretaban como designio divino. Así fue que hace alrededor de mil años, un obispo, de lo que ahora es Bélgica, determinó que existen 56 maneras diferentes en las que tres dados pueden caer *a condición que uno esté interesado sólo en el resultado global y no en qué hace cada dado*. Asignó una virtud a cada una de estas posibilidades y cada pecador tenía que concentrarse durante cierto tiempo en la virtud que correspondía a su tirada de dados.

(a) Encuentre el número de maneras en que tres dados pueden caer con el mismo número.
(b) Obtenga el número de maneras en que dos de los tres dados pueden caer con el mismo número de puntos y el tercero caiga con un número diferente.
(c) Determine el número de maneras en que los tres dados pueden caer con números diferentes.
(d) Use los resultados de las partes (a), (b) y (c) para verificar los cálculos del obispo de que hay en total 56 posibilidades.

1.29 Si la NCAA tiene solicitudes de seis universidades para ser el anfitrión de los campeonatos interuniversitarios de tenis en 1998 y 1999, ¿de cuántas maneras pueden seleccionar al anfitrión para estos campeonatos
(a) si ambos no se van a celebrar en la misma universidad;
(b) si ambos pueden realizarse en la misma universidad?

1.30 Las cinco finalistas del concurso señorita Universo son las representantes de Argentina, Bélgica, Estados Unidos, Japón y Noruega. ¿De cuántas maneras pueden los jueces escoger a
(a) la ganadora y la primera suplente;
(b) la ganadora, la primera y la segunda suplentes?

1.31 En una elección primaria, hay cuatro candidatos para el puesto de alcalde, cinco para tesorero de la ciudad, y dos candidatos para procurador.
(a) ¿De cuántas maneras puede un votante marcar su boleta para elegir a los tres funcionarios?
(b) ¿De cuántas maneras puede una persona votar si ejerce su elección de no votar por un candidato para alguno o todos estos puestos?

1.32 Una prueba de elección múltiple consta de 15 preguntas, cada una permite una elección entre tres alternativas. ¿De cuántas maneras diferentes puede una estudiante marcar sus respuestas a estas preguntas?

1.33 El precio de un recorrido turístico por Europa incluye cuatro sitios qué visitar que deben seleccionarse a partir de 10 ciudades. ¿De cuántas maneras diferentes se puede planear tal viaje
(a) si es importante el orden de las paradas intermedias;
(b) si no es importante el orden de las paradas intermedias?

1.34 ¿De cuántas maneras puede un director de televisión programar los seis diferentes anuncios de un patrocinador, durante los seis espacios de tiempo asignado para anuncios, durante un "especial" de una hora?

1.35 ¿De cuántas maneras puede el director de televisión del ejercicio 1.34 asignar los seis espacios de tiempo para anuncios si el patrocinador tiene tres anuncios diferentes, cada uno de los cuales se puede mostrar dos veces?

1.36 ¿De cuántas maneras puede el director de televisión del ejercicio 1.34 cubrir los seis espacios de tiempo para anuncios si el patrocinador tiene dos anuncios diferentes, cada uno de los cuales se puede mostrar tres veces?

1.37 ¿De cuántas maneras se pueden formar en línea cinco personas para subir a un autobús? ¿De cuántas maneras se pueden formar en línea si dos de las personas se rehúsan a hacerlo una detrás de la otra?

1.38 ¿De cuántas maneras pueden ocho personas formar un círculo para un baile folklórico?

1.39 ¿Cuántas permutaciones hay de las letras en la palabra
(a) "great" (grandioso);
(b) "greet" (saludo)?

1.40 ¿Cuántas permutaciones hay de las letras de la palabra "statistics" (estadística)? ¿Cuántas empiezan y terminan con la letra *s*?

1.41 Un equipo colegial juega 10 partidos de fútbol durante una temporada. ¿De cuántas maneras puede terminar la temporada con cinco juegos ganados, cuatro perdidos y un empate?

1.42 Si ocho personas están reunidas para comer, ¿de cuántas maneras diferentes tres de ellas pueden ordenar pollo, cuatro ordenar carne y una ordenar langosta?

1.43 En el ejemplo 1.4 demostramos que una prueba de falso o verdadero que consta de 20 preguntas, se puede marcar de 1,048,576 maneras diferentes. ¿De cuántas maneras se puede marcar cada pregunta con falso o verdadero de modo que
(a) 7 estén correctas y 13 incorrectas;
(b) 10 estén correctas y 10 incorrectas;
(c) al menos 17 estén correctas?

1.44 Entre los siete candidatos para dos vacantes en el consejo de una ciudad hay tres hombres y cuatro mujeres. ¿De cuántas maneras se pueden cubrir estas vacantes
(a) con dos candidatos cualquiera de los siete;
(b) con dos de las cuatro mujeres;
(c) con uno de los hombres y una de las mujeres?

1.45 Entre 10 aparatos de televisión de un embarque, hay tres que están defectuosos. ¿De cuántas maneras puede un hotel comprar cuatro de estos aparatos y recibir al menos dos de los aparatos defectuosos?

1.46 Ms. Jones tiene cuatro faldas, siete blusas y tres suéteres. ¿De cuántas maneras puede escoger dos de las faldas, tres de las blusas y uno de los suéteres para llevar en un viaje?

1.47 ¿Cuántas manos de bridge diferentes pueden contener cinco espadas, tres diamantes, tres tréboles y dos corazones?

1.48 Encuentre el número de maneras en que una A, tres B, dos C y una F se pueden distribuir entre siete estudiantes que toman un curso de estadística.

1.49 Una coleccionista de arte, dueña de 10 pinturas de artistas famosos, está preparando su testamento. ¿De cuántas maneras diferentes puede dejar estas pinturas a sus tres herederos?

1.50 Un aficionado al béisbol tiene dos boletos para seis juegos diferentes en el estadio de los Cachorros de Chicago. Si tiene cinco amigos a quienes les gusta el béisbol, ¿de cuántas maneras diferentes puede invitar a uno de ellos a cada uno de los seis juegos?

1.51 Al final del día, una pastelería da todo lo que no se vendió a centros de acopio de comida para los necesitados. Si, al final de un día dado, le quedan 12 pasteles de manzana, ¿de cuántas maneras diferentes puede distribuir estos pasteles entre seis centros de comida para los necesitados?

1.52 Con respecto al ejercicio 1.51, ¿de cuántas maneras diferentes puede la pastelería distribuir los 12 pasteles de manzana si cada uno de los centros de comida va a recibir al menos un pastel?

1.53 Un viernes por la mañana, la tienda de artículos para profesionales de un club de tenis tiene 14 latas idénticas de pelotas para tenis. Si para el domingo en la noche se han vendido todas y sólo nos interesa cuántas se vendieron cada día, ¿de cuántas maneras diferentes se pudieron haber vendido las pelotas de tenis el viernes, el sábado y el domingo?

1.54 Vuelva a realizar el ejercicio 1.53, dado que al menos dos de las latas de pelotas para tenis se vendieron en cada uno de los tres días.

REFERENCIAS

Entre los pocos libros sobre la historia de la estadística se cuentan

WALKER, H. M., *Studies in the History of Statistical Method*. Baltimore: The Williams & Wilkins Company, 1929,

WESTERGAARD, H., *Contributions to the History of Statistics*. Londres: P. S. King & Son, 1932,

y las publicaciones más recientes

KENDALL, M. G., and PLACKETT, R. L., eds., *Studies in the History of Statistics and Probability*, Vol. II. Nueva York: Macmillan Publishing Co., Inc., 1977,

PEARSON, E. S., and KENDALL, M. G., eds., *Studies in the History of Statistics and Probability*. Darien, Conn.: Hafner Publishing Co., Inc., 1970,

PORTER, T. M., *The Rise of Statistical Thinking, 1820–1900*. Princeton, N.J.: Princeton University Press, 1986,

STIGLER, S. M., *The History of Statistics*. Cambridge, Mass.: Harvard University Press, 1986.

Una amplia variedad de material sobre métodos combinatorios se puede encontrar en

COHEN, D. A., *Basic Techniques of Combinatorial Theory*. Nueva York: John Wiley & Sons, Inc., 1978,

EISEN, M., *Elementary Combinatorial Analysis*. Nueva York: Gordon and Breach, Science Publishers, Inc., 1970,

FELLER, W., *An Introduction to Probability Theory and Its Applications*, Vol. I, 3rd ed. Nueva York: John Wiley & Sons, Inc., 1968,

NIVEN, J., *Mathematics of Choice*. Nueva York: Random House, Inc., 1965,

ROBERTS, F. S., *Applied Combinatorics*. Upper Saddle River, N.J.: Prentice Hall, 1984,

y en

WHITWORTH, W. A., *Choice and Chance*, 5th ed. New York: Hafner Publishing Co., Inc., 1959,
que se ha convertido en un clásico en este campo.

Se pueden encontrar tratamientos más avanzados en

BECKENBACH, E. F., ed., *Applied Combinatorial Mathematics*. Nueva York: John Wiley & Sons, Inc., 1964,

DAVID, F. N., and BARTON, D. E., *Combinatorial Chance*. Nueva York: Hafner Publishing Co., Inc., 1962,

y

RIORDAN, J., *An Introduction to Combinatorial Analysis*. Nueva York: John Wiley & Sons, Inc., 1958.

CAPÍTULO

2

Probabilidad

2.1 INTRODUCCIÓN

Históricamente, la forma más antigua de definir probabilidades, el **concepto clásico de probabilidad**, se aplica cuando todos los resultados posibles son igualmente probables, como es presumiblemente el caso en la mayoría de los juegos de azar. Podemos entonces decir que *si hay N posibilidades igualmente probables, de las cuales una debe ocurrir y n se consideran favorables, o como un "acierto," entonces la probabilidad de un "acierto" está dada por la razón* $\frac{n}{N}$.

EJEMPLO 2.1

¿Cuál es la probabilidad de sacar un as de una baraja ordinaria de 52 cartas de juego?

Solución

Puesto que hay $n = 4$ ases entre las $N = 52$ cartas, la probabilidad de sacar un as es $\frac{4}{52} = \frac{1}{13}$. (Se supone, por supuesto, que cada carta tiene la misma oportunidad de salir.) ▲

Aunque las posibilidades igualmente probables se encuentran principalmente en los juegos de azar, el concepto clásico de probabilidad también se aplica en una gran variedad de situaciones donde se usan los dispositivos de juego para hacer selecciones aleatorias (cuando se asigna al azar el espacio de la oficina para los asistentes de enseñanza, cuando algunas familias de un municipio se escogen de manera que cada una tenga la misma oportunidad de ser incluida en un estudio, muestra de cuando las partes de una máquina se escogen para inspección de tal manera que cada parte producida tenga la misma oportunidad de ser seleccionada, y así sucesivamente).

Una deficiencia importante del concepto clásico de probabilidad es su aplicación limitada, pues hay muchas situaciones en que las posibilidades que se presentan no pueden considerarse igualmente probables. Éste sería el caso, por ejemplo, si nos interesara la cuestión de si lloverá cierto día, si nos interesara el resultado de una elección, o si nos concierne la mejoría de una persona enferma.

Entre los diversos conceptos de probabilidad, el más ampliamente sostenido es, la **interpretación de frecuencia** de acuerdo a la cual *la probabilidad de un evento (resultado o suceso)* es la proporción de las veces en que eventos de la misma clase ocurrirán en un largo plazo. Si decimos que la probabilidad de que un jet de Los Ángeles a San Francisco llegue a tiempo es de 0.84, queremos decir (de acuerdo con la interpretación de frecuencia) que tales vuelos llegarán a tiempo 84% de las veces. En forma similar, si el servicio meteorológico predice que hay 30% de posibilidades de lluvia (esto es, una probabilidad de 0.30), esto significa que bajo las mismas condiciones del clima lloverá 30%. En términos más generales, decimos que un evento tiene una probabilidad de, por ejemplo 0.90, en el mismo sentido en que podríamos decir que nuestro automóvil arrancará en clima frío 90% del tiempo. No podemos garantizar lo que sucederá en una ocasión en particular (el automóvil puede encender ahora y después tal vez no) pero si llevamos registros durante un largo periodo, debemos encontrarnos con que la proporción de "aciertos" es muy cercana a 0.90.

Un punto de vista alternativo, que actualmente se ve favorecido, consiste en interpretar las probabilidades como **evaluaciones personales** o **subjetivas**. Tales probabilidades expresan la fuerza de lo que creemos respecto a las incertidumbres que están en juego, y se aplican especialmente cuando hay poca o ninguna evidencia directa, así que no hay más opción que considerar evidencia colateral (indirecta), "suposiciones educadas", y quizá la intuición u otros factores subjetivos.

El enfoque a la probabilidad que usaremos en este capítulo es el **enfoque axiomático**, en el que las probabilidades se definen como "objetos matemáticos" que se comportan de acuerdo a ciertas reglas bien definidas. Entonces, cualquiera de los conceptos o interpretaciones de probabilidad anteriores se puede usar en aplicaciones en tanto sea congruente con estas reglas.

2.2 ESPACIOS MUESTRALES

Puesto que todas las probabilidades pertenecen a la ocurrencia o no ocurrencia de eventos, expliquemos primero el significado de *evento* y de los términos relacionados *experimento, resultado* y *espacio muestral.*

En estadística se acostumbra denominar **experimento** a cualquier proceso de observación o medición. En este sentido, un experimento puede consistir en el sencillo proceso de verificar si un interruptor está encendido o apagado; puede consistir en contar las imperfecciones en un pedazo de tela; o puede consistir en el tan complicado proceso de medir la masa de un electrón. Los productos de un experimento, ya sean lecturas de instrumentos, cuentas, respuestas "sí" o "no", o valores obtenidos mediante cálculos extensos, se conocen como **resultados** del experimento.

Al conjunto de todos los posibles resultados de un experimento se le conoce como el **espacio muestral** y suele representarse con la letra S. Cada resultado de un espacio muestral se llama **elemento** del espacio muestral o simplemente un **punto de la muestra**. Si el espacio muestral tiene un número finito de elementos, podemos enumerar los elementos en la notación usual de conjuntos; por ejemplo, el espacio muestral de los posibles resultados de tirar una moneda se puede escribir como

$$S = \{\text{H}, \text{T}\}$$

donde H y T representan cara y cruz. Los espacios muestrales con un número de elementos grande, o infinito, se describen mejor con un enunciado o una regla; por ejemplo, si los posibles resultados de un experimento son el conjunto de automóviles equipados con radios de banda civil, el espacio muestral se puede escribir

$$S = \{x | x \text{ es un automóvil con radio de BC}\}$$

Esto se lee "S es el conjunto de toda x tal que x es un automóvil con radio de BC". De la misma forma, si S es el conjunto de los enteros positivos impares, escribimos

$$S = \{2k + 1 | k = 0, 1, 2, \ldots\}$$

La manera en que formulemos el espacio muestral en una situación dada dependerá del problema que se tenga. Si un experimento consiste en lanzar una vez un dado y nos interesara qué lado queda hacia arriba, usaríamos el espacio muestral

$$S_1 = \{1, 2, 3, 4, 5, 6\}$$

Sin embargo si sólo nos interesara que la cara que queda hacia arriba sea par o impar, usaríamos el espacio de muestreo

$$S_2 = \{\text{par}, \text{impar}\}$$

Esto demuestra que bien se pueden usar diferentes espacios muestrales para describir un experimento. En general, *es deseable usar espacios muestrales cuyos elementos no se puedan dividir (partir o separar) en clases de resultados más primitivos o más elementales.* En otras palabras, *es preferible que un elemento de un espacio muestral no represente dos o más resultados que son distinguibles en alguna manera.* Así, en la ilustración precedente S_1 sería preferible a S_2.

EJEMPLO 2.2

Describa un espacio muestral que sea apropiado para un experimento en el que tiramos un par de dados, uno rojo y uno verde.

Solución

El espacio muestral que proporciona la mayor información consiste en los 36 puntos dados por

$$S_1 = \{(x, y) | x = 1, 2, \ldots, 6; y = 1, 2, \ldots, 6\}$$

donde x representa el número en que cayó el dado rojo y y representa el número del dado verde. Un segundo espacio muestral, adecuado para la mayoría de los propósitos (aunque menos deseable en general ya que proporciona menos información), está dado por

$$S_2 = \{2, 3, 4, \ldots, 12\}$$

donde los elementos son los totales de los números en que cayeron los dos dados.

▲

Los espacios muestrales se suelen clasificar de acuerdo al número de elementos que contienen. En el ejemplo anterior los espacios muestrales S_1 y S_2 contenían un número **finito** de elementos; pero si se lanza una moneda hasta que aparezca una cara por primera vez, esto podría suceder en el primer lanzamiento, el segundo lanzamiento, el tercer lanzamiento, el cuarto lanzamiento, ..., y hay infinitamente muchas posibilidades. Para este experimento obtenemos el espacio muestral

$$S = \{\text{H}, \text{TH}, \text{TTH}, \text{TTTH}, \text{TTTTH}, \ldots\}$$

con una secuencia interminable de elementos. Pero aun en este caso el número de elementos se puede igualar uno a uno con los números enteros y en este sentido se dice que el espacio muestral es **contable**. Si el espacio muestral contiene un número finito de elementos o un número infinito aunque contable de elementos, se dice que es **discreto.**

Los resultados de algunos experimentos no son ni finitos ni contablemente infinitos. Tal es el caso, por ejemplo, cuando uno realiza una investigación para determinar la distancia a la que cierta marca de automóviles viajará, con una ruta de prueba prescrita, con 5 litros de gasolina. Si suponemos que la distancia es una variable que puede medirse con cualquier grado de exactitud deseado, hay una infinidad de posibilidades (distancias) que no se pueden igualar uno a uno con los números enteros. También, si queremos medir la cantidad de tiempo que dos sustancias químicas tardan en reaccionar, las cantidades que forman el espacio muestral son infinitas en número y no son contables. Así, los espacios muestrales no necesitan ser discretos. Si un espacio muestral consiste en un continuo, tal como los puntos de un segmento de línea o todos los puntos de un plano, se dice que es **continuo**. Los espacios muestrales continuos surgen en la práctica siempre que los resultados de los experimentos son mediciones de propiedades físicas, como temperatura, velocidad, presión, longitud, ..., que se miden con escalas continuas.

2.3 EVENTOS

En muchos problemas nos interesan resultados que no son dados directamente por un elemento específico de un espacio muestral.

EJEMPLO 2.3

Con respecto al primer espacio muestral S_1 en la página 27, describa el evento A en que el número de puntos obtenidos con el dado sea divisible entre 3.

Solución

Entre 1, 2, 3, 4, 5 y 6, sólo 3 y 6 son divisibles entre 3. Por consiguiente, A está representado por el subconjunto $\{3, 6\}$ del espacio muestral S_1. ▲

EJEMPLO 2.4

Respecto al espacio muestral S_1 del ejemplo 2.2, describa el evento B en que el número de puntos obtenidos con el par de dados es 7.

Solución

Entre las 36 posibilidades, sólo (1, 6), (2, 5), (3, 4), (4, 3), (5, 2) y (6, 1) dan un total de 7. Así, escribimos

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

Observe que en la figura 2.1 el evento de que caiga un total de 7 con los dos dados se representa con el conjunto de puntos dentro de la región limitada por la línea punteada. ▲

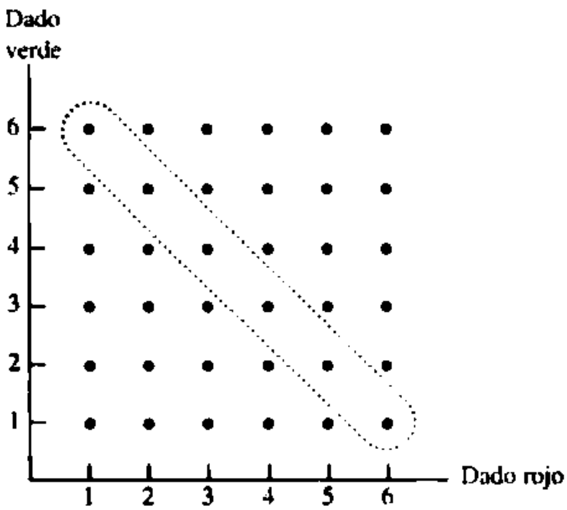

Figura 2.1 Obtener un total de 7 con un par de dados.

De la misma manera, cualquier evento (desenlace o resultado) se puede identificar con un grupo de puntos, los que constituyen un subconjunto de un espacio muestral apropiado. Tal subconjunto consta de todos los elementos de un espacio muestral para los cuales el evento ocurre y en probabilidad y estadística identificamos el subconjunto con el evento. Así, por definición, un **evento** es un subconjunto de un espacio muestral.

EJEMPLO 2.5

Si alguien dispara a un blanco tres veces y sólo nos interesa si cada disparo da o no en el blanco, describa un espacio muestral apropiado, los elementos del espacio muestral que constituyen el evento M que la persona no acertará en el blanco tres veces seguidas, y los elementos del evento N que la persona acertará una vez y fallará en dos ocasiones.

Solución

Si dejamos que 0 y 1 representen una falla y un acierto respectivamente, las ocho posibilidades (0, 0, 0), (1, 0, 0), (0, 1, 0), (0, 0, 1), (1, 1, 0), (1, 0, 1), (0, 1, 1) y (1, 1, 1) se pueden mostrar como en la figura 2.2. Así, se puede ver que

$$M = \{(0, 0, 0)\}$$

y

$$N = \{(1, 0, 0), (0, 1, 0), (0, 0, 1)\} \qquad \blacktriangle$$

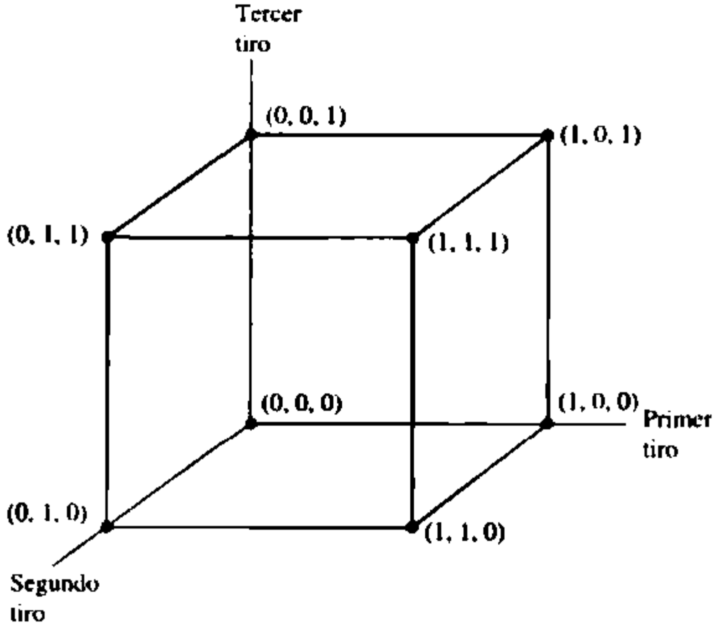

Figura 2.2 Espacio muestral para el ejemplo 2.5.

EJEMPLO 2.6

Construya un espacio muestral para la duración de la vida útil de cierto componente electrónico e indique el subconjunto que represente el evento F de que el componente falle antes del final del sexto año.

Solución

Si t es la duración de la vida útil del componente en años, el espacio muestral se puede escribir $S = \{t|t \geqq 0\}$, y el subconjunto $F = \{t|0 \leqq t < 6\}$ es el evento de que el componente falle antes del final del sexto año. ▲

De acuerdo a nuestra definición, cualquier evento es un subconjunto de un espacio muestral apropiado, pero debe observarse que la recíproca no es necesariamente verdad. Para espacios muestrales discretos, todos los subconjuntos son eventos, pero en el caso continuo algunos conjuntos de puntos más bien oscuros se deben excluir por razones matemáticas. Esto se examina con más detalle en alguno de los textos avanzados que se encuentran entre las referencias al final de este capítulo, pero no es importante por lo que concierne al trabajo en este libro.

En muchos problemas de probabilidad nos interesan eventos que en realidad son combinaciones de dos o más eventos, formados al tomar **uniones, intersecciones** y **complementos**. Aunque el lector seguramente estará familiarizado con estos términos, revisemos brevemente que si A y B son dos subconjuntos cualquiera de un espacio muestral S, su unión $A \cup B$ es el subconjunto de S que contiene todos los elementos que están en A, en B o en ambos; su intersección $A \cap B$ es el subconjunto de S que contiene todos los elementos que están tanto en A como en B; y el complemento A' de A es el subconjunto de S que contiene todos los elementos de S que no están en A. En los ejercicios 2.1 a 2.4 se puede encontrar algunas de las reglas que controlan la formación de uniones, intersecciones y complementos.

A menudo se describen los espacios muestrales y los eventos, particularmente las relaciones entre eventos, por medio de **diagramas de Venn**, en los cuales el espacio muestral se representa con un rectángulo, en tanto que los eventos se denotan con regiones dentro del rectángulo, usualmente con círculos o partes de círculos. Por ejemplo, las regiones sombreadas en los cuatro diagramas de Venn de la figura 2.3 representan, respectivamente, al evento A, al complemento del evento A, a la unión de los eventos A y B, y a la intersección de los eventos A y B. Cuando estamos trabajando con tres eventos, por lo común dibujamos los círculos como en la figura 2.4. Aquí, las regiones se numeran del 1 al 8 para facilitan la referencia.

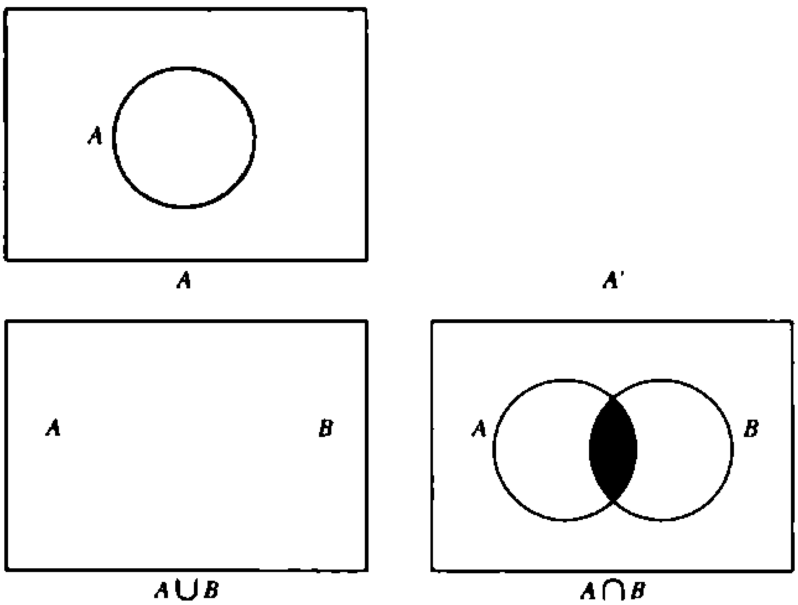

Figura 2.3 Diagramas de Venn.

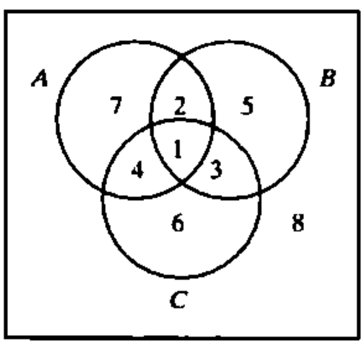

Figura 2.4 Diagrama de Venn.

A y B son mutuamente excluyentes A está contenido en B

Figura 2.5 Diagramas que muestran relaciones especiales entre eventos.

Para indicar las relaciones especiales entre eventos, a veces dibujamos diagramas como los de la figura 2.5. Aquí, el de la izquierda sirve para indicar que los eventos A y B son **mutuamente excluyentes**; esto es, que los dos conjuntos no tienen elementos en común (o que los dos eventos no pueden ocurrir al mismo tiempo). Cuando A y B son mutuamente excluyentes, escribimos $A \cap B = \emptyset$, donde $\emptyset$ denota el **conjunto vacío**, el cual no tiene elemento alguno. El diagrama de la derecha sirve para indicar que A está contenido **en** B, y simbólicamente lo expresamos con $A \subset B$.

EJERCICIOS

2.1 Use diagramas de Venn para verificar que
- (a) $(A \cup B) \cup C$ es el mismo evento que $A \cup (B \cup C)$;
- (b) $A \cap (B \cup C)$ es el mismo evento que $(A \cap B) \cup (A \cap C)$;
- (c) $A \cup (B \cap C)$ es el mismo evento que $(A \cup B) \cap (A \cup C)$.

2.2 Con diagramas de Venn verifique las dos **leyes de De Morgan**:
- (a) $(A \cap B)' = A' \cup B'$;
- (b) $(A \cup B)' = A' \cap B'$.

2.3 Use diagramas de Venn para verificar que si A está contenida en B, entonces $A \cap B = A$ y $A \cap B' = \emptyset$.

2.4 Con diagramas de Venn verifique que
- (a) $(A \cap B) \cup (A \cap B') = A$;
- (b) $(A \cap B) \cup (A \cap B') \cup (A' \cap B) = A \cup B$;
- (c) $A \cup (A' \cap B) = A \cup B$.

APLICACIONES

2.5 Si $S = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$, $A = \{1, 3, 5, 7\}$, $B = \{6, 7, 8, 9\}$, $C = \{2, 4, 8\}$, y $D = \{1, 5, 9\}$, liste los elementos de los subconjuntos de S que corresponden a los siguientes eventos:
- (a) $A' \cap B$; (b) $(A' \cap B) \cap C$; (c) $B' \cup C$;
- (d) $(B' \cup C) \cap D$; (e) $A' \cap C$; (f) $(A' \cap C) \cap D$.

2.6 Una empresa de electrónica planea construir un laboratorio de investigación en el sur de California y la dirección debe elegir entre locales en Los Ángeles, en San Diego, en Long Beach, en Pasadena, en Santa Bárbara, en Anaheim, en Santa Mónica, y en Westwood. Si A representa el evento de que escogerán un local en San Diego o Santa Bárbara, B representa el evento de que se decidirán por un sitio en San Diego o Long Beach, C representa el evento de que lo eligirán en Santa Bárbara o Anaheim, y D representa el evento de que lo escogerán en Los Ángeles o Santa Bárbara, haga una lista de los elementos de cada uno de los siguientes subconjuntos del espacio muestral, que consiste en ocho selecciones de local:

(a) A'; (b) D'; (c) $C \cap D$;
(d) $B \cap C$; (e) $B \cup C$; (f) $A \cup B$;
(g) $C \cup D$; (h) $(B \cup C)'$; (i) $B' \cap C'$.

2.7 Entre los ocho automóviles que un vendedor tiene en su sala de exhibición, el automóvil 1 es nuevo, tiene aire acondicionado, dirección hidráulica y asientos de cubo; el vehículo 2, tiene un año de uso, tiene aire acondicionado, pero no tiene ni dirección hidráulica ni asientos de cubo; el automóvil 3, tiene dos años de uso, tiene aire acondicionado y dirección hidráulica, pero no tiene asientos de cubo; la unidad 4 tiene tres años de uso, tiene aire acondicionado pero no tiene ni dirección hidráulica ni asientos de cubo; el vehículo 5 es nuevo, no tiene aire acondicionado, ni dirección hidráulica ni asientos de cubo; el automóvil 6 tiene un año de uso, tiene dirección hidráulica, pero no tiene ni aire acondicionado ni asientos de cubo; el vehículo 7 tiene dos años de uso, no tiene aire acondicionado, ni dirección hidráulica ni asientos de cubo; y la unidad 8 tiene tres años de uso, no tiene aire acondicionado, pero tiene dirección hidráulica así como asientos de cubo. Si un cliente compra uno de estos automóviles y el evento de que compre un vehículo nuevo, por ejemplo, se representa con el conjunto {automóvil 1, automóvil 5}, indique en forma similar los conjuntos que representan los eventos de que

(a) se decida por un automóvil sin aire acondicionado;
(b) escoja una unidad sin dirección hidráulica;
(c) escoja un vehículo con asientos de cubo;
(d) escoja un automóvil que tenga dos o tres años de uso.

2.8 Con respecto al ejercicio 2.7, enuncie con palabras qué clase de automóvil escogerá el cliente, si su elección está dada por

(a) el complemento del conjunto del inciso (a);
(b) la unión de los conjuntos de los incisos (b) y (c);
(c) la intersección de los conjuntos de los incisos (c) y (d);
(d) la intersección de los incisos (b) y (c) de este ejercicio.

2.9 Si la señora Brown compra una de las casas anunciadas para su venta en un diario de Seattle (en un domingo dado), T es el evento de que la casa tiene tres o más baños, U es el evento de que tiene una chimenea, V es el evento de que cuesta más de $100,000, y W es el evento de que es nueva, describa (con palabras) cada uno de los siguientes eventos:

(a) T'; (b) U'; (c) V';
(d) W'; (e) $T \cap U$; (f) $T \cap V$;
(g) $U' \cap V$; (h) $V \cup W$; (i) $V' \cup W$;
(j) $T \cup U$; (k) $T \cup V$; (l) $V \cap W$.

2.10 Un hotel recreativo tiene dos camionetas, que usa para trasladar a sus huéspedes del hotel al aeropuerto y viceversa. Si la más grande de las dos camionetas puede llevar cinco pasajeros y la más pequeña puede llevar cuatro pasajeros, el punto (0, 3) representa el evento de que en un momento dado la camioneta más grande está vacía, en tanto que las más pequeña tiene tres pasajeros, el punto (4, 2) representa el evento de que en un momento dado la camioneta más grande tiene cuatro pasajeros en tanto que las más pequeña tiene dos pasajeros, ..., dibuje una figura que muestre los 30 puntos del espacio muestral correspondiente. También, si E representa el evento de que al menos una de las camionetas está vacía, F representa el evento de que juntas llevan dos, cuatro o seis pasajeros, y G representa el evento que cada una lleva el mismo número de pasajeros, enumere los puntos del espacio muestral que corresponde a cada uno de los siguientes eventos:

(a) E;	(b) F;	(c) G;
(d) $E \cup F$;	(e) $E \cap F$;	(f) $F \cup G$;
(g) $E \cup F'$;	(h) $E \cap G'$;	(i) $F' \cap E'$.

2.11 Se lanza una moneda al aire una vez. Entonces, si cae cara, se tira un dado una vez; si cae cruz, el dado se tira dos veces más. Utilice la notación en la que (H, 2), por ejemplo, denota el evento de que la moneda cae cara y entonces el dado cae en 2, y (T, T, T) denota el evento de que la moneda cae cruz tres veces seguidas, para enumerar

(a) los 10 elementos del espacio muestral S;
(b) los elementos de S que corresponden al evento A de que caiga exactamente una cara;
(c) los elementos de S que corresponden al evento B de que caiga al menos dos veces cruz o un número mayor que 4.

2.12 Un juego electrónico contiene tres componentes dispuestos en el circuito en serie paralelo de la figura 2.6. En un momento dado cualquiera, cada componente puede estar en operación o no, y el juego funcionará sólo si hay un circuito ininterrumpido de P a Q. Sea A el evento de que el juego funcionará; sea B el evento de que el juego funcionará aunque el componente x no esté en operación; y sea C el evento de que el juego funcionará aunque el componente y no esté en operación. Use la notación en la cual (0, 0, 1), por ejemplo, denota que el componente z está en operación pero los componentes x y y no lo están, y

(a) enumere los elementos del espacio muestral S y también los elementos de S que corresponden a los eventos A, B y C;
(b) determine qué pares de eventos, A y B, A y C o B y C, son mutuamente excluyentes.

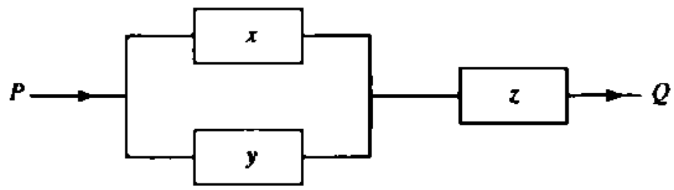

Figura 2.6 Diagrama para el ejercicio 2.12.

2.13 Un experimento consiste en tirar un dado hasta que aparezca un 3. Describa el espacio muestral y determine

(a) cuántos elementos del espacio muestral corresponden al evento de que el 3 aparezca en el $k$ésimo tiro del dado;

(b) cuántos elementos del espacio muestral corresponden al evento de que el 3 no caerá después del $k$ésimo tiro del dado.

2.14 Si $S = \{x|0 < x < 10\}$, $M = \{x|3 < x \leq 8\}$, y $N = \{x|5 < x < 10\}$, encuentre

(a) $M \cup N$; (b) $M \cap N$;

(c) $M \cap N'$; (d) $M' \cup N$.

2.15 Exprese simbólicamente el espacio muestral S que consiste en todos los puntos (x, y) sobre o dentro de una circunferencia de radio 3 centrado en el punto $(2, -3)$.

2.16 En la figura 2.7, L es el evento de que una conductora tenga seguro de responsabilidad civil y C es el evento de que tenga seguro contra accidentes. Exprese con palabras qué eventos están representados en las regiones 1, 2, 3 y 4.

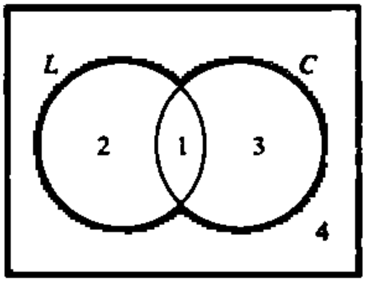

Figura 2.7 Diagrama de Venn para el ejercicio 2.16.

2.17 Con respecto al ejercicio 2.16 y la figura 2.7, ¿qué eventos están representados en

(a) las regiones 1 y 2 juntas;

(b) las regiones 2 y 4 juntas;

(c) las regiones 1, 2 y 3 juntas;

(d) las regiones 2, 3 y 4 juntas?

2.18 En la figura 2.8, E, T, y N son los eventos de que un automóvil en un taller necesite una reparación mayor del motor, reparaciones en la transmisión o neumáticos nuevos. Con palabras exprese los eventos representados en

Figura 2.8 Diagrama de Venn para el ejercicio 2.18.

(a) región 1;
(b) región 3;
(c) región 7;
(d) regiones 1 y 4 juntas;
(e) regiones 2 y 5 juntas;
(f) regiones 3, 5, 6 y 8 juntas.

2.19 Con respecto al ejercicio 2.18 y la figura 2.8, enumere la región o combinación de regiones que represente los eventos de que un automóvil en el taller necesite
(a) reparaciones de la transmisión, pero no reparación mayor del motor ni neumáticos nuevos;
(b) una reparación mayor del motor y reparaciones de la transmisión;
(c) reparaciones de la transmisión o neumáticos nuevos, pero no una reparación mayor del motor;
(d) neumáticos nuevos.

2.20 En un grupo de 200 estudiantes universitarios, 138 están inscritos en un curso de psicología, 115 están en un curso de sociología, y 91 están inscritos en ambos. ¿Cuántos de estos estudiantes no están inscritos en ninguno de los cursos? (*Sugerencia:* Dibuje un diagrama de Venn apropiado y anote los números asociados con las diversas regiones.)

2.21 Una organización de investigación de mercado afirma que, de 500 compradores entrevistados, 308 compran regularmente el producto X, 266 el producto Y, 103 compran regularmente ambos, y 59 no compran ninguno en forma regular. Utilice un diagrama de Venn y anote el número de compradores asociado con las diversas regiones para verificar si el resultado de este estudio debe ponerse en duda.

2.22 De 120 visitantes de Disneylandia, 74 permanecieron en el parque por lo menos 3 horas, 86 gastaron al menos \$20, 64 tomaron el paseo del Matterhorn, 60 se quedaron al menos 3 horas y gastaron por lo menos \$20, 52 permanecieron cuando menos 3 horas y tomaron el paseo del Matterhorn, 54 gastaron al menos \$20 y tomaron el paseo del Matterhorn y 48 permanecieron 3 horas cuando menos, gastaron al menos \$20 y tomaron el paseo del Matterhorn. Dibuje un diagrama de Venn con tres círculos (como el de la figura 2.4) y anote los números asociados con las diversas regiones, encuentre cuántos de los 120 visitantes a Disneylandia
(a) permanecieron en el parque al menos 3 horas, gastaron \$20 por lo menos, pero no tomaron el paseo del Matterhorn;
(b) tomaron el paseo del Matterhorn, pero permanecieron en el parque menos de 3 horas y gastaron menos de \$20;
(c) permanecieron en el parque menos de 3 horas, gastaron al menos \$20, pero no tomaron el paseo del Matterhorn.

2.4 LA PROBABILIDAD DE UN EVENTO

Para formular los postulados de probabilidad, seguiremos la práctica de denotar los eventos mediante letras mayúsculas, y escribiremos la probabilidad del evento A como $P(A)$, la probabilidad del evento B como $P(B)$, y así sucesivamente. Como antes, denotaremos el conjunto de todos los resultados posibles, el espacio muestral, con la letra S.

Las probabilidades son los valores de una función de conjunto, también conocida como **medida de probabilidad**, ya que, como veremos, esta función asigna números reales a los diversos subconjuntos de un espacio muestral S. Tal como los formularemos aquí, los postulados de probabilidad se aplican sólo cuando el espacio muestral S es discreto.

POSTULADO 1 La probabilidad de un evento es un número real no negativo; esto es, $P(A) \geqq 0$ para cualquier subconjunto A de S.

POSTULADO 2 $P(S) = 1$.

POSTULADO 3 Si $A_1, A_2, A_3, \ldots$, es una secuencia finita o infinita de eventos mutuamente excluyentes de S, entonces

$$P(A_1 \cup A_2 \cup A_3 \cup \cdots) = P(A_1) + P(A_2) + P(A_3) + \cdots$$

Los postulados *per se* no requieren demostración, pero si se va a aplicar la teoría resultante debemos demostrar que se satisfacen los postulados cuando damos a las probabilidades un significado "real". Ilustremos esto aquí, en relación con la interpretación de frecuencia; la relación entre los postulados y el concepto clásico de probabilidad se examinará en la página 41, mientras que la relación entre los postulados y las probabilidades subjetivas se deja al lector para su examen en los ejercicios 2.34 y 2.56.

Puesto que las proporciones siempre son positivas o cero, el primer postulado está en completo acuerdo con la interpretación de frecuencia. El segundo postulado enuncia indirectamente que la certidumbre está identificada con una probabilidad de 1; después de todo, siempre se supone que debe ocurrir una de las posibilidades en S, y es a este evento cierto que asignamos una probabilidad de 1. Hasta donde concierne a la interpretación de frecuencia, una probabilidad de 1 implica que el evento en cuestión ocurrirá 100% de las veces o, en otras palabras, que ocurrirá con certeza.

Tomando el tercer postulado en el caso más simple, que es para dos eventos mutuamente excluyentes A_1 y A_2, se puede ver fácilmente que se cumple por la interpretación de frecuencia. Si un evento ocurre, digamos, 28% de las veces, otro evento ocurre 39% de las veces, y ambos eventos no pueden ocurrir al mismo tiempo (es decir, son mutuamente excluyentes), entonces uno o el otro ocurrirá $28 + 39 = 67\%$ de las veces. Así, el tercer postulado se cumple, y el mismo tipo de argumento se aplica cuando hay más de dos eventos mutuamente excluyentes.

Antes de que estudiemos algunas de las consecuencias inmediatas de los postulados de probabilidad, subrayemos el punto que los tres postulados no nos dicen cómo asignar probabilidades a los eventos; ellos meramente r ringen las maneras en que se puede hacer.

EJEMPLO 2.7

Un experimento tiene cuatro resultados posibles, A, B, C y D, que son mutuamente excluyentes. Explique por qué las siguientes asignaciones de probabilidades no están permitidas:

(a) $P(A) = 0.12, P(B) = 0.63, P(C) = 0.45, P(D) = -0.20$;

(b) $P(A) = \frac{9}{120}, P(B) = \frac{45}{120}, P(C) = \frac{27}{120}, P(D) = \frac{46}{120}$.

Solución

(a) $P(D) = -0.20$ viola el postulado 1;

(b) $P(S) = P(A \cup B \cup C \cup D) = \frac{9}{120} + \frac{45}{120} + \frac{27}{120} + \frac{46}{120} = \frac{127}{120} \neq 1$,

y esto viola el postulado 2. ▲

Por supuesto, en la práctica real las probabilidades se asignan con base en la experiencia pasada, sobre la base de un análisis cuidadoso de las condiciones subyacentes, sobre la base de juicios subjetivos, o sobre la base de suposiciones —algunas veces la suposición de que todos los resultados posibles son equiprobables.

Para asignar una medida de probabilidad a un espacio muestral, no es necesario especificar la probabilidad para cada subconjunto posible. Esto es afortunado, pues un espacio muestral con tan pocas como 20 resultados posibles ya tiene $2^{20} = 1{,}048{,}576$ subconjuntos [la fórmula general resulta directamente del inciso (a) del ejercicio 1.14], y el número de subconjuntos crece muy rápidamente cuando hay 50 resultados posibles, 100 resultados posibles o más. En vez de enumerar las probabilidades de todos los subconjuntos posibles, a menudo listamos las probabilidades de los resultados individuales, o puntos muestra de S, y entonces hacemos uso del teorema siguiente.

> **TEOREMA 2.1** Si A es un evento en un espacio muestral discreto S, entonces $P(A)$ es igual a la suma de las probabilidades de los resultados individuales que abarcan A.

Demostración. Sean $O_1, O_2, O_3, \ldots$, la secuencia finita o infinita de resultados que abarcan el evento A. Así

$$A = O_1 \cup O_2 \cup O_3 \cdots$$

y puesto que los resultados individuales, las O, son mutuamente excluyentes, el tercer postulado de la probabilidad nos da

$$P(A) = P(O_1) + P(O_2) + P(O_3) + \cdots$$

Esto completa la prueba. ▼

Para usar este teorema, debemos poder asignar probabilidades a los resultados individuales de los experimentos. Los ejemplos siguientes ilustran cómo se hace esto en algunas situaciones especiales.

EJEMPLO 2.8

Si lanzamos dos veces una moneda balanceada, ¿cuál es la probabilidad de sacar al menos una cara?

Solución

El espacio muestral es $S = \{HH, HT, TH, TT\}$, donde H y T denotan cara y cruz. Puesto que suponemos que la moneda está balanceada, estos resultados son igualmente posibles y asignamos a cada punto muestra la probabilidad de $\frac{1}{4}$. Denotemos con A el evento que sacamos al menos una cara, obtenemos $A = \{\text{HH}, \text{HT}, \text{TH}\}$ y

$$\begin{aligned} P(A) &= P(\text{HH}) + P(\text{HT}) + P(\text{TH}) \\ &= \frac{1}{4} + \frac{1}{4} + \frac{1}{4} \\ &= \frac{3}{4} \quad \blacktriangle \end{aligned}$$

EJEMPLO 2.9

Un dado está arreglado de manera que cada número impar tiene el doble de probabilidad de ocurrir que un número par. Encuentre $P(G)$, donde G es el evento que un número mayor que 3 ocurra en un solo tiro del dado.

Solución

El espacio muestral es $S = \{1, 2, 3, 4, 5, 6\}$. Por tanto, si asignamos la probabilidad w a cada número par y la probabilidad $2w$ a cada número impar, encontramos que $2w + w + 2w + w + 2w + w = 9w = 1$ de acuerdo al postulado 2. Se deduce que $w = \frac{1}{9}$ y

$$P(G) = \frac{1}{9} + \frac{2}{9} + \frac{1}{9} = \frac{4}{9} \qquad \blacktriangle$$

Si un espacio muestral es contablemente infinito, se tendrán que asignar las probabilidades a los resultados individuales mediante una regla matemática, preferentemente mediante una fórmula o una ecuación.

EJEMPLO 2.10

Si para un experimento dado, $O_1, O_2, O_3, \ldots$, es una secuencia infinita de resultados, verificar que

$$P(O_i) = \left(\frac{1}{2}\right)^i \qquad \text{para } i = 1, 2, 3, \ldots$$

es, realmente, una medida de probabilidad.

Solución

Puesto que las probabilidades son todas positivas, queda por demostrar que $P(S) = 1$. Al obtener

$$P(S) = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \cdots$$

y mediante la fórmula para la suma de términos de una progresión geométrica infinita, encontramos que

$$P(S) = \frac{\frac{1}{2}}{1 - \frac{1}{2}} = 1 \qquad \blacktriangle$$

En relación con el ejemplo precedente, la palabra "suma" en el teorema 2.1 tendrá que ser interpretada para que incluya el valor de una serie infinita.

Como veremos en el capítulo 5, la medida de probabilidad del ejemplo 2.10 sería apropiada, por ejemplo, si O_i es el evento de que una persona que lanza una moneda balanceada obtendrá una cruz por primera vez en el $i$ésimo lanzamiento de la moneda. Así, la probabilidad de que la primera cruz venga en el tercer, cuarto o quinto lanzamiento de la moneda es

$$\left(\frac{1}{2}\right)^3 + \left(\frac{1}{2}\right)^4 + \left(\frac{1}{2}\right)^5 = \frac{7}{32}$$

y la probabilidad de que la primera cruz salga en un lanzamiento de número impar es

$$\left(\frac{1}{2}\right)^1 + \left(\frac{1}{2}\right)^3 + \left(\frac{1}{2}\right)^5 + \cdots = \frac{\frac{1}{2}}{1 - \frac{1}{4}} = \frac{2}{3}$$

Aquí otra vez usamos la fórmula para la suma de los términos de una progresión geométrica infinita.

Si un experimento es tal que podemos suponer probabilidades iguales para todos los puntos muestra, como fue el caso en el ejemplo 2.8, podemos tomar ventaja del siguiente caso especial del teorema 2.1.

TEOREMA 2.2 Si un experimento puede resultar en cualquiera de N resultados diferentes igualmente probables, y si n de estos resultados juntos constituyen el evento A, entonces la probabilidad del evento A es

$$P(A) = \frac{n}{N}$$

Demostración. Representemos los resultados individuales en S con O_1, O_2 ,..., O_N cada uno con una probabilidad $\frac{1}{N}$. Si A es la unión de n de estos resultados mutuamente excluyentes, y no importa cuales, entonces

$$\begin{aligned} P(A) &= P(O_1 \cup O_2 \cup \cdots \cup O_n) \\ &= P(O_1) + P(O_2) + \cdots + P(O_n) \\ &= \underbrace{\frac{1}{N} + \frac{1}{N} + \cdots + \frac{1}{N}}_{n \text{ términos}} \\ &= \frac{n}{N} \qquad \blacktriangledown \end{aligned}$$

Observe que la fórmula $P(A) = \frac{n}{N}$ del teorema 2.2 es idéntica con la del concepto clásico de probabilidad (ver página 25). En verdad, lo que hemos demostrado aquí es que el concepto clásico de probabilidad es congruente con los postulados de probabilidad —resulta de los postulados en el caso especial donde los resultados individuales son todos equiprobables.

EJEMPLO 2.11

Se dice que una mano de póker de cinco cartas repartidas de un baraja de 52 cartas de juego es un "full" si consiste en tres de un mismo valor y un par. Si todas las manos de cinco cartas son igualmente probables, ¿cuál es la probabilidad de que le den un "full"?

Solución

El número de maneras en que nos pueden dar un "full" en particular, digamos tres reyes y dos ases, es $\binom{4}{3}\binom{4}{2}$. Puesto que hay 13 maneras de seleccionar el valor de la carta para las tres del mismo valor y para cada una de éstas hay 12 maneras de seleccionar el valor de la carta para el par, en total hay

$$n = 13 \cdot 12 \cdot \binom{4}{3}\binom{4}{2}$$

diferentes "fulles". También el número total de manos de póker de cinco cartas es

$$N = \binom{52}{5}$$

y resulta de acuerdo al teorema 2.2 que la probabilidad de obtener un "full" es

$$P(A) = \frac{n}{N} = \frac{13 \cdot 12\binom{4}{3}\binom{4}{2}}{\binom{52}{5}} = 0.0014 \qquad \blacktriangle$$

2.5 ALGUNAS REGLAS DE PROBABILIDAD

Basados en los tres postulados de probabilidad, podemos derivar muchas otras reglas que tienen aplicaciones importantes. Entre ellas, los cuatro teoremas siguientes son consecuencia inmediata de los postulados.

TEOREMA 2.3 Si A y A' son eventos complementarios en un espacio muestral S, entonces

$$P(A') = 1 - P(A)$$

Demostración. En el segundo y tercer pasos de la prueba que sigue, usamos la definición de complemento, de acuerdo a la cual A y A' son mutuamente excluyentes y $A \cup A' = S$. Así, escribimos

$$\begin{aligned} 1 &= P(S) && \text{(por el postulado 2)} \\ &= P(A \cup A') \\ &= P(A) + P(A') && \text{(por el postulado 3)} \end{aligned}$$

y de ahí resulta que $P(A') = 1 - P(A)$. ▼

En relación con la interpretación de frecuencia, este resultado implica que si un evento ocurre, digamos, 37% de las veces, entonces no ocurre 63% de las veces.

TEOREMA 2.4 $P(\emptyset) = 0$ para cualquier espacio muestral S.

Demostración. Puesto que S y $\emptyset$ son mutuamente excluyentes y $S \cup \emptyset = S$ de acuerdo con la definición del conjunto vacío $\emptyset$, resulta que

$$\begin{aligned} P(S) &= P(S \cup \emptyset) \\ &= P(S) + P(\emptyset) && \text{(por el postulado 3)} \end{aligned}$$

y, por tanto, que $P(\emptyset) = 0$. ▼

Es importante señalar que no resulta necesariamente que si $P(A) = 0$ entonces $A = \emptyset$. En la práctica, a menudo asignamos la probabilidad 0 a eventos que, en términos coloquiales, no sucederían en un millón de años. Por ejemplo, hay el ejemplo clásico que le asignamos una probabilidad de 0, al evento de un mono con una máquina de escribir, escribirá *La República* de Platón palabra por palabra sin un error. Como veremos en los capítulos 3 y 6, es relevante el hecho que $P(A) = 0$ no implica que $A = \emptyset$ es pertinente, especialmente, en el caso continuo.

TEOREMA 2.5 Si A y B son eventos en un espacio muestral S y $A \subset B$, entonces $P(A) \leqq P(B)$.

Demostración. Puesto que $A \subset B$, podemos escribir

$$B = A \cup (A' \cap B)$$

como se puede verificar fácilmente mediante un diagrama de Venn. Entonces, puesto que A y $A' \cap B$ son mutuamente excluyentes, obtenemos

$$\begin{aligned} P(B) &= P(A) + P(A' \cap B) && \text{(por el postulado 3)} \\ &\geqq P(A) && \text{(por el postulado 1)} \end{aligned}$$ ▼

En palabras, este teorema enuncia que si A es un subconjunto de B, entonces $P(A)$ no puede ser mayor que $P(B)$. Por ejemplo, la probabilidad de sacar un corazón de un baraja ordinaria de 52 cartas de juego no puede ser mayor que la probabilidad de sacar una carta roja. En verdad, la probabilidad es $\frac{1}{4}$, comparada con $\frac{1}{2}$.

TEOREMA 2.6 $0 \leqq P(A) \leqq 1$ para cualquier evento A.

Demostración. Usando el teorema 2.5 y el hecho que $\emptyset \subset A \subset S$ para cualquier evento A en S, tenemos

$$P(\emptyset) \leqq P(A) \leqq P(S)$$

Entonces, $P(\emptyset) = 0$ y $P(S) = 1$ nos lleva al resultado que

$$0 \leqq P(A) \leqq 1$$ ▼

A veces nos referimos al tercer postulado de probabilidad como la **regla especial de adición**; es especial en el sentido que los eventos $A_1, A_2, A_3, \ldots$, deben ser todos mutuamente excluyentes. Para dos eventos cualquiera A y B, existe la **regla general de adición**:

TEOREMA 2.7 Si A y B son dos eventos en el espacio muestral S, entonces

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

Demostración. Si asignamos las probabilidades a, b, y c a los eventos mutuamente excluyentes $A \cap B$, $A \cap B'$ y $A' \cap B$ como en el diagrama de Venn de la figura 2.9, encontramos que

$$\begin{aligned} P(A \cup B) &= a + b + c \\ &= (a + b) + (c + a) - a \\ &= P(A) + P(B) - P(A \cap B) \quad \blacktriangledown \end{aligned}$$

Figura 2.9 Diagrama de Venn para la demostración del teorema 2.7.

EJEMPLO 2.12

En un zona metropolitana grande, las probabilidades son 0.86, 0.35 y 0.29 de que una familia (escogida aleatoriamente para una encuesta de muestreo) tenga un aparato de televisión a color, un aparato de televisión en blanco y negro, o ambas clases de aparatos respectivamente. ¿Cuál es la probabilidad de que una familia posea cualquiera de los dos o ambas clases de aparatos?

Solución

Si A es el evento de que una familia en esta zona metropolitana tenga un aparato de televisión a color y B es el evento de que tiene un aparato blanco y negro, tenemos $P(A) = 0.86$, $P(B) = 0.35$ y $P(A \cap B) = 0.29$; al sustituir en la fórmula del teorema 2.7 nos da

$$\begin{aligned} P(A \cup B) &= 0.86 + 0.35 - 0.29 \\ &= 0.92 \quad \blacktriangle \end{aligned}$$

EJEMPLO 2.13

Cerca de cierta salida de la carretera I-17, las probabilidades son 0.23 y 0.24, de que un camión parado en un retén tendrá frenos defectuosos o neumáticos muy gastados. También, la probabilidad es 0.38 de que un camión parado en el retén tendrá frenos defectuosos y/o neumáticos muy gastados. ¿Cuál es la probabilidad de que un camión parado en este retén tendrá los frenos defectuosos así como los neumáticos muy gastados?

Solución

Si B es el evento que un camión parado en el retén tendrá frenos defectuosos y T es el evento de que tendrá neumáticos muy gastados, tenemos $P(B) = 0.23$, $P(T) = 0.24$ y $P(B \cup T) = 0.38$; al sustituir en la fórmula del teorema 2.7 nos da

$$0.38 = 0.23 + 0.24 - P(B \cap T)$$

Al resolver para $P(B \cap T)$, obtenemos

$$P(B \cap T) = 0.23 + 0.24 - 0.38 = 0.09 \quad \blacktriangle$$

Al usar repetidamente la fórmula del teorema 2.7, podemos generalizar esta regla de adición de manera que se aplique a cualquier número de eventos. Por ejemplo, para tres eventos obtenemos

TEOREMA 2.8 Si A, B y C son tres eventos cualquiera en el espacio muestral S, entonces

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

Demostración. Al escribir $A \cup B \cup C$ como $A \cup (B \cup C)$ y al aplicar la fórmula del teorema 2.7 dos veces, una vez para $P[A \cup (B \cup C)]$ y una vez para $P(B \cup C)$, obtenemos

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

Entonces, usamos la ley distributiva que se pidió al lector que verificara en el inciso (b) del ejercicio 2.1, encontramos que

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

y por tanto que

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

(En el ejercicio 2.30 se pedirá al lector que dé una demostración alternativa de este teorema, basado en el método usado en el texto para demostrar el teorema 2.7.)

EJEMPLO 2.14

Si una persona acude con su dentista, supongamos que la probabilidad de que le limpie la dentadura es 0.44, la probabilidad de que le tape una caries es 0.24, la probabilidad de que se le extraiga un diente es 0.21, la probabilidad de que se le limpie la dentadura y le tape una caries es 0.08, la probabilidad de que le limpie la dentadura y le extraiga un diente es 0.11, la probabilidad de que le tape una caries y le saque un diente es 0.07, y la probabilidad de que le limpie la dentadura, le tape una caries y le saque un diente es 0.03. ¿Cuál es la probabilidad de que a una persona que acude con su dentista se le haga por lo menos una de estas cosas?

Solución

Si C es el evento que a la persona se le limpie la dentadura, F es el evento que se le tape una caries, y E es el evento de que se le saque un diente, se nos da $P(C) = 0.44$, $P(F) = 0.24$, $P(E) = 0.21$, $P(C \cap F) = 0.08$, $P(C \cap E) = 0.11$, $P(F \cap E) = 0.07$ y $P(C \cap F \cap E) = 0.03$, y la sustitución en la fórmula del teorema 2.8 nos da

$$\begin{aligned} P(C \cup F \cup E) &= 0.44 + 0.24 + 0.21 - 0.08 - 0.11 - 0.07 + 0.03 \\ &= 0.66 \quad \blacktriangle \end{aligned}$$

EJERCICIOS

2.23 Use las partes (a) y (b) del ejercicio 2.4 para demostrar que

(a) $P(A) \geqq P(A \cap B)$;

(b) $P(A) \leqq P(A \cup B)$.

2.24 Con referencia a la figura 2.9, verificar que

$$P(A \cap B') = P(A) - P(A \cap B)$$

2.25 Con referencia a la figura 2.9 y haciendo que $P(A' \cap B') = d$, verificar que

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

2.26 El evento que "A o B pero no ambos" ocurrirá se puede escribir como

$$(A \cap B') \cup (A' \cap B)$$

Exprese la probabilidad de este evento en términos de $P(A)$, $P(B)$ y $P(A \cap B)$.

2.27 Use la fórmula del teorema 2.7 para demostrar que

(a) $P(A \cap B) \leqq P(A) + P(B)$;

(b) $P(A \cap B) \geqq P(A) + P(B) - 1$.

2.28 Use el diagrama de Venn de la figura 2.10 con las probabilidades a, b, c, d, e, f y g asignadas a $A \cap B \cap C$, $A \cap B \cap C', \ldots,$ y $A \cap B' \cap C'$ para mostrar que si $P(A) = P(B) = P(C) = 1$, entonces $P(A \cap B \cap C) = 1$. (*Sugerencia*: Empiece con el argumento que puesto que $P(A) = 1$, se concluye que $e = c = f = 0$.)

2.29 Dé una demostración alternativa del teorema 2.7 usando las relaciones $A \cup B = A \cup (A' \cap B)$ y $B = (A \cap B) \cup (A' \cap B)$.

2.30 Use el diagrama de Venn de la figura 2.10 y el método por el que se demostró el teorema 2.7 para probar el teorema 2.8

Figura 2.10 Diagrama para los ejercicios 2.28, 2.30 y 2.31.

2.31 Repetir el método de demostración usado en el ejercicio 2.30 para demostrar que

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

(*Sugerencia*: Con respecto al diagrama de Venn de la figura 2.10 divida cada una de las ocho regiones en dos partes, designando a una estar dentro de D y la otra fuera de D y sean $a, b, c, d, e, f, g, h, i, j, k, l, m, n, o$ y p las probabilidades asociadas con las 16 regiones resultantes.

2.32 Demuestre por inducción que

$$P(E_1 \cup E_2 \cup \cdots \cup E_n) \leq \sum_{i=1}^{n} P(E_i)$$

para cualquier secuencia finita de eventos $E_1, E_2, \ldots,$ y E_n.

2.33 La **ventaja** de que un evento ocurrirá está dada por la razón de la probabilidad de que el evento ocurra a la probabilidad de que no ocurra, siempre que ninguna de las probabilidades sea cero. La ventaja generalmente se indica en términos de enteros positivos que no tienen un factor en común. Mostrar que si la ventaja es a a b de que un evento ocurrirá, su probabilidad es

$$p = \frac{a}{a+b}$$

2.34 Se pueden determinar las probabilidades subjetivas al exponer a las personas a situaciones donde se corren riesgos y al encontrar la ventaja a la cual considerarían justo apostar al resultado. La ventaja entonces se convierte en probabilidades por medio de la fórmula del ejercicio 2.33. Por ejemplo, si una persona siente que la ventaja de 3 a 2 es ventaja justa de que una empresa comercial tendrá éxito (o que sería justo apostar \$30 contra \$20 a que tendrá éxito), la probabilidad es $\frac{3}{3+2} = 0.6$ de que la empresa comercial tendrá éxito. Demuestre que si las probabilidades subjetivas se determinan de esta manera, satisfacen

(a) el postulado 1 en la página 37;

(b) el postulado 2.

Véase también el ejercicio 2.56.

APLICACIONES

2.35 Un experimento tiene cinco resultados posibles, A, B, C, D, y E, que son mutuamente excluyentes. Verifique si las asignaciones de probabilidades siguientes son permisibles y explique sus respuestas

(a) $P(A) = 0.20$, $P(B) = 0.20$, $P(C) = 0.20$, $P(D) = 0.20$ y $P(E) = 0.20$;

(b) $P(A) = 0.21$, $P(B) = 0.26$, $P(C) = 0.58$, $P(D) = 0.01$ y $P(E) = 0.06$;

(c) $P(A) = 0.18$, $P(B) = 0.19$, $P(C) = 0.20$, $P(D) = 0.21$ y $P(E) = 0.22$;

(d) $P(A) = 0.10$, $P(B) = 0.30$, $P(C) = 0.10$, $P(D) = 0.60$ y $P(E) = -0.10$;

(e) $P(A) = 0.23$, $P(B) = 0.12$, $P(C) = 0.05$, $P(D) = 0.50$ y $P(E) = 0.08$.

2.36 Si A y B son mutuamente excluyentes, $P(A) = 0.37$ y $P(B) = 0.44$, encontrar

(a) $P(A')$; (b) $P(B')$; (c) $P(A \cup B)$;

(d) $P(A \cap B)$; (e) $P(A \cap B')$; (f) $P(A' \cap B')$.

2.37 Explique por qué hay un error en cada una de las siguientes declaraciones:

(a) La probabilidad de que Jean apruebe el examen de la barra de abogados es 0.66 y la probabilidad de que no lo pase es -0.34.

(b) La probabilidad de que el equipo de casa gane un juego de fútbol venidero es 0.77, la probabilidad de que se empate el juego es 0.08, y la probabilidad de que gane o empate el juego es 0.95.

(c) Las probabilidades de que una secretaria cometa 0, 1, 2, 3, 4, 5 o más errores al mecanografiar un informe son, respectivamente, 0.12, 0.25, 0.36, 0.14, 0.09 y 0.07.

(d) Las probabilidades de que un banco reciba 0, 1, 2, 3 o más cheques malos en un día dado son, respectivamente, 0.08, 0.21, 0.29 y 0.40.

2.38 Supongamos que a cada uno de los 30 puntos del espacio muestral del ejercicio 2.10 se le asigna la probabilidad $\frac{1}{30}$. Encuentre las probabilidades de que en un momento dado

(a) al menos una de las camionetas esté vacía;

(b) cada una de las camionetas transporte el mismo número de pasajeros;

(c) la camioneta más grande transporte más pasajeros que la camioneta más pequeña;

(d) juntas transporten al menos seis pasajeros.

2.39 Las probabilidades de que la facilidad de darle servicio a una nueva máquina de rayos X se clasifique como muy difícil, difícil, promedio, fácil o muy fácil son, respectivamente, 0.12, 0.17, 0.34, 0.29 y 0.08. Encuentre las probabilidades de que la facilidad de darle servicio a la máquina se clasifique

(a) difícil o muy difícil;

(b) ni muy difícil ni muy fácil;

(c) promedio o peor;

(d) promedio o mejor.

2.40 Un departamento de policía necesita neumáticos nuevos para sus carros patrullas y las probabilidades son 0.15, 0.24, 0.03, 0.28, 0.22 y 0.08 respectivamente que comprará neumáticos Uniroyal, neumáticos Goodyear, neumáticos Michelin, neumáticos General, neumáticos Goodrich o neumáticos Armstrong. Encuentre las probabilidades de que comprará

(a) neumáticos Goodyear o Goodrich;

(b) neumáticos Uniroyal, Michelin, o Goodrich;

(c) neumáticos Michelin, o Armstrong;

(d) neumáticos Uniroyal , Michelin, General, o Goodrich.

2.41 Un sombrero contiene veinte papeletas blancas numeradas del 1 al 20, diez papeletas rojas numeradas del 1 al 10, cuarenta papeletas amarillas numeradas del 1 al 40, y diez papeletas azules numeradas del 1 al 10. Si estas papeletas se mezclan muy bien para que cada una tenga la misma probabilidad de salir, encuentre las probabilidades de sacar una papeleta que sea:

(a) azul o blanca;

(b) numerada 1, 2, 3, 4 o 5;

(c) roja o amarilla y numerada 1, 2, 3 o 4;

(d) numerada, 5, 15, 25 o 35;

(e) blanca y con número mayor que 12 o amarilla y con número mayor que 26.

2.42 Cuatro candidatos están buscando una vacante en un consejo escolar. Si A tiene el doble de posibilidades que B de ser elegido, y a A y a B se le dan las mismas oportunidades de ser electos, mientras que C tiene el doble de posibilidades que D de ser electo, ¿cuáles son las probabilidades de que

(a) C gane;

(b) A no gane.

2.43 Dos cartas se extraen aleatoriamente de una baraja de 52 cartas de juego. Encuentre la probabilidad de que ambas cartas sean mayores que 3 y menores que 8.

2.44 En un juego de póker, cinco cartas se reparten aleatoriamente de una baraja ordinaria de 52 cartas de juego. Encuentre las probabilidades de sacar

(a) dos pares (dos valores cualquiera distintos que ocurren dos veces exactamente)

(b) cuatro de una clase (cuatro cartas con el mismo valor).

2.45 En un juego de Yahtzee, se tiran simultáneamente cinco dados balanceados. Encuentre las probabilidades de sacar

(a) dos pares;

(b) tres de una clase;

(c) un "full" (tres de una clase y un par);

(d) cuatro de una clase.

2.46 De los 78 doctores del personal de un hospital, 64 tienen seguro contra tratamiento erróneo, 36 son cirujanos y 34 de los cirujanos tienen seguro contra tratamiento erróneo. Si uno de estos doctores se escoge al azar para representar al personal del hospital en una convención de la A.M.A. (esto es, cada doctor tiene una probabilidad de $\frac{1}{78}$ de ser seleccionado), ¿cuál es la probabilidad de que el seleccionado no sea un cirujano y no tenga seguro contra tratamiento erróneo?

2.47 Explique sobre la base de las diversas reglas de los ejercicios 2.23 a 2.27 por qué hay un error en cada uno de los siguientes enunciados:

(a) La probabilidad de que llueva es 0.67, y la probabilidad de que llueva o nieve es de 0.55.

(b) La probabilidad de que una estudiante obtenga una calificación aprobatoria en inglés es 0.82, y la probabilidad de que obtenga una calificación aprobatoria en inglés y francés es 0.86.

(c) La probabilidad de que una persona que visite el zoológico de San Diego vea las jirafas es 0.72, la probabilidad de que vea los osos es de 0.84 y la probabilidad de que vea ambos es 0.52.

2.48 Dado $P(A) = 0.59$, $P(B) = 0.30$ y $P(A \cap B) = 0.21$, encontrar

(a) $P(A \cup B)$; (b) $P(A \cap B')$;

(c) $P(A' \cup B')$; (d) $P(A' \cap B')$.

2.49 Para parejas casadas que viven en cierto suburbio, la probabilidad de que el marido vote en una elección del consejo escolar es de 0.21, la probabilidad de que la esposa vote es de 0.28, y la probabilidad de que ambos voten es de 0.15. ¿Cuál es la probabilidad de que al menos uno de ellos vote?

2.50 Una profesora de biología tiene dos asistentes graduados que la ayudan con su investigación. La probabilidad de que el mayor de los dos asistentes se ausente en un día dado es 0.08, la probabilidad de que el más joven de los dos se ausente en un día dado es 0.05 y la probabilidad de que ambos se ausenten en un día dado es 0.02. Encuentre las probabilidades de que

(a) cualquiera o ambos de los asistentes graduados esté ausente en cualquier día dado;

(b) al menos uno de los dos asistentes graduados no esté ausente en cualquier día dado;

(c) sólo uno de los dos asistentes graduados esté ausente en cualquier día dado.

2.51 En el Roanoke College se sabe que $\frac{1}{3}$ de los estudiantes no viven en el campus. También se sabe que $\frac{5}{9}$ de los estudiantes son oriundos del estado de Virginia y que $\frac{3}{4}$ de los estudiantes son de fuera del estado o viven en el campus. ¿Cuál es la probabilidad de que un estudiante seleccionado aleatoriamente del Roanoke College sea de fuera del estado y viva en el campus?

2.52 Suponga que si una persona visita Disneylandia, la probabilidad de que vaya al Crucero de la Jungla es 0.74, la probabilidad de que se suba al Monorriel es 0.70, la probabilidad de que tome el paseo del Matterhorn es 0.62, la probabilidad de que vaya al Crucero de la Jungla y se suba al Monorriel es 0.52, la probabilidad de que vaya al Crucero de la Jungla así como tome el paseo del Matterhorn es 0.46, la probabilidad de que se suba al Monorriel y tome el paseo del Matterhorn es 0.44, y la probabilidad de que haga estas tres cosas es 0.34. ¿Cuál es la probabilidad de que una persona que visite Disneylandia hará al menos una de estas tres cosas?

2.53 Supongamos que si una persona viaja a Europa por primera vez, la probabilidad de que visite Londres es 0.70, la probabilidad de que visite París es 0.64, la probabilidad de que visite Roma es 0.58, la probabilidad de que visite Amsterdam es 0.58, la probabilidad de que visite Londres y París es 0.45, la probabilidad de que visite Londres y Roma es 0.42, la probabilidad de que visite Londres y Amsterdam es 0.41, la probabilidad de que visite París y Roma es 0.35, la probabilidad de que visite París y Amsterdam es 0.39, la probabilidad de que visite Roma y Amsterdam es 0.32, la probabilidad de que visite Londres, París y Roma es 0.23, la probabilidad de que visite Londres, París y Amsterdam es 0.26, la probabilidad de que visite Londres, Roma y Amsterdam es 0.21, la probabilidad de que visite París, Roma y Amsterdam es 0.20, y la probabilidad de que visite todas estas cuatro ciudades es 0.12. ¿Cuál es la probabilidad de que una persona que viaja a Europa por primera vez visite al menos una de estas cuatro ciudades. (*Sugerencia*: Use la fórmula del ejercicio 2.31.)

2.54 Use la fórmula del ejercicio 2.33 para convertir cada una de las siguientes ventajas en probabilidades:

(a) Si se escogen aleatoriamente tres huevos de una caja de 12 huevos de los cuales tres están rotos, la ventaja es de 34 a 21 de que al menos uno de ellos estará roto.

(b) Si una persona tiene ocho billetes de \$1, cinco billetes de \$5, y un billete de \$20, y aleatoriamente selecciona tres de ellos, la ventaja es de 11 a 2 de que no todos serán billetes de \$1.

(c) Si disponemos arbitrariamente las letras de la palabra "nest", la ventaja es de 5 a 1 de que no obtendremos una palabra que signifique algo en inglés.

2.55 Use la definición de "ventaja" dado en el ejercicio 2.33 para convertir cada una de las siguientes probabilidades a ventaja:

(a) La probabilidad de que el último dígito de las placas de circulación de un automóvil sea un 2, 3, 4, 5, 6 o 7 es de $\frac{6}{10}$.

(b) La probabilidad de sacar al menos dos caras en cuatro lanzamientos de una moneda balanceada es $\frac{11}{16}$.

(c) La probabilidad de tirar "7 u 11" con un par de dados balanceados es $\frac{2}{9}$.

2.56 Si las probabilidades subjetivas se determinan por el método sugerido por el ejercicio 2.34, el tercer postulado de probabilidad puede no satisfacerse. Sin embargo, los proponentes del concepto de probabilidad subjetiva generalmente imponen este postulado como un **criterio de congruencia**; en otras palabras, consideran las probabilidades subjetivas que no satisfacen el postulado como incongruentes.

(a) La directora de una secundaria piensa que la ventaja es 7 a 5 contra ella para obtener un aumento de \$1,000 y 11 a 1 contra ella de obtener un aumento de \$2,000. Además, piensa que obtener uno de estos aumentos o el otro es una apuesta pareja. Examine la congruencia de las probabilidades subjetivas correspondientes.

(b) Cuando a un funcionario de un partido político le preguntan sobre su futuro político, responde que la ventaja es 2 a 1 a que no se postulará para la Cámara de Representantes, y 4 a 1 a que no se postulará para el Senado. Además, siente que la ventaja es 7 a 5 a que se postulará para una o el otro. ¿Son congruentes las probabilidades correspondientes?

2.57 Hay dos Porsches en una carrera de automóviles en Italia, y un reportero cree que la ventaja contra que ganen es de 3 a 1 y de 5 a 3. Para ser congruente (véase el ejercicio 2.56), ¿qué ventaja debe asignar el reportero al evento que uno de los dos automóviles gane?

2.58 Si hacemos x = al número de puntos en las caras de arriba cuando se lanzan un par de dados, entonces podemos usar el espacio de muestreo S_2 del ejemplo 2.2 para describir los resultados del experimento.

(a) Encuentre la probabilidad de cada resultado en S_2.

(b) Verifique que la suma de estas probabilidades es 1.

2.59 Use un programa de computadora que pueda generar enteros aleatoriamente en el intervalo (0, 10) con iguales probabilidades, genere 1,000 de tales enteros y use la interpretación de frecuencia para estimar la probabilidad de que un entero escogido aleatoriamente tenga un valor menor que 1.

2.60 Use el método del ejercicio 2.59, genere un segundo conjunto de enteros aleatorios en (0, 10). Estime la probabilidad de que A: un entero seleccionado aleatoriamente del primer conjunto sea menor que 1 ó B: un entero seleccionado aleatoriamente del segundo conjunto sea menor que 1.

(a) use la interpretación de frecuencia de las probabilidades;

(b) use el teorema 2.7 y $P(A \cap B) = 0.25$.

2.6 PROBABILIDAD CONDICIONAL

Cuando las probabilidades se citan sin especificar el espacio muestral pueden aparecer dificultades con facilidad. Por ejemplo, si preguntamos la probabilidad de que un abogado gane más de \$50,000 al año, bien podría ser que obtuviéramos varias respuestas diferentes y todas podrían ser correctas. Una de ellas podría aplicarse a todos los graduados de facultades de leyes, otra podría aplicarse a todas las personas con licencia para practicar la profesión legal, una tercera podría aplicarse a todas aquellas dedicadas activamente a la práctica de la profesión legal, y así sucesivamente. Puesto que la elección de un espacio muestral (esto es, el conjunto de todas las posibilidades bajo consideración) no es siempre evidente por sí mismo, a menudo ayuda a usar el símbolo $P(A|S)$ para denotar la **probabilidad condicional** del evento A relativa al espacio muestral S o, como también la llamamos, "la probabilidad de A dado S". El símbolo $P(A|S)$ hace explícito que nos estamos refiriendo a un espacio muestral particular S, y es preferible que la notación abreviada $P(A)$ a menos que la elección tácita de S se entienda claramente. También es preferible cuando queremos referirnos a varios espacios muestrales en el mismo ejemplo. Si A es el evento de que una persona gane más de \$50,000 al año, G es el evento de que una persona sea un graduado de la facultad de leyes, L es el evento de que una persona tenga licencia para practicar la profesión legal, y E es el evento de que una persona esté dedicada activamente a la práctica de la profesión legal, entonces $P(A|G)$ es la probabilidad de que un graduado de la facultad de leyes gane más de \$50,000 al año, $P(A|L)$ es la probabilidad de que una persona con licencia para practicar la profesión legal gane más de \$50,000 al año y $P(A|E)$ es la probabilidad de que una persona dedicada activamente a la práctica de la profesión legal gane más de \$50,000 al año.

En los siguientes ejemplos se ilustran algunas ideas relacionadas con las probabilidades condicionales.

EJEMPLO 2.15

Una organización de investigación de los consumidores ha estudiado los servicios con garantía proporcionados por las 50 agencias de automóviles nuevos en una cierta ciudad, y en la tabla siguiente se resumen sus hallazgos.

	Buen servicio de garantía	*Mal servicio de garantía*
En operación por 10 años o más	16	4
En operación menos de 10 años	10	20

Si una persona selecciona aleatoriamente una de estas agencias de automóviles nuevos, ¿cuál es la probabilidad de que seleccione una que proporciona buen servicio de garantía? También, si una persona selecciona aleatoriamente una de las agencias que han operado por 10 años o más, ¿cuál es la probabilidad de que seleccione una agencia que proporciona buen servicio de garantía?

Solución

Por "aleatoriamente" queremos decir que, en cada caso, todas las selecciones son igualmente probables y podemos por tanto usar la fórmula del teorema 2.2. Si G denota la selección de la agencia que proporciona buen servicio de garantía, y si $n(G)$ denota el número de elementos en G y $n(S)$ el número de elementos en el espacio muestral completo, obtenemos

$$P(G) = \frac{n(G)}{n(S)} = \frac{16 + 10}{50} = 0.52$$

Esto contesta la primera pregunta.

Para la segunda pregunta, nos circunscribimos al espacio muestral reducido que consta de la primera línea de la tabla, esto es, las $16 + 4 = 20$ agencias que han estado operando por 10 años o más. De éstas, 16 proporcionan buen servicio de garantía y obtenemos

$$P(G|T) = \frac{16}{20} = 0.80$$

donde T denota la selección de una agencia que ha operado por 10 años o más. Esto responde la segunda pregunta y, como se debió haber esperado, $P(G|T)$ es considerablemente más alta que $P(G)$. ▲

Puesto que el numerador de $P(G|T)$ es $n(T \cap G) = 16$ en el ejemplo precedente, el número de agencias que han operado por 10 años o más y brindan buen servicio de garantía y el denominador es $n(T)$, el número de agencias que han operado 10 años o más, podemos escribir con símbolos

$$P(G|T) = \frac{n(T \cap G)}{n(T)}$$

Entonces, si dividimos el numerador y el denominador entre $n(S)$, el número total de agencias de automóviles nuevos en la ciudad dada, obtenemos

$$P(G|T) = \frac{\frac{n(T \cap G)}{n(S)}}{\frac{n(T)}{n(S)}} = \frac{P(T \cap G)}{P(T)}$$

y así, hemos expresado la probabilidad condicional $P(G|T)$ en términos de dos probabilidades definida para todo el espacio muestral S.

Generalizando de lo anterior, definamos ahora la probabilidad condicional.

DEFINICIÓN 2.1 Si A y B son dos eventos cualquiera en un espacio muestral S y $P(A) \neq 0$, la **probabilidad condicional** de B dado A es

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

EJEMPLO 2.16

Con respecto al ejemplo 2.15, ¿cuál es la probabilidad de que una de las agencias que ha operado menos de 10 años proporcionará buen servicio de garantía?

Solución

Puesto que $P(T' \cap G) = \frac{10}{50} = 0.20$ y $P(T') = \frac{10 + 20}{50} = 0.60$, la sustitución en la fórmula nos da

$$P(G|T') = \frac{P(T' \cap G)}{P(T')} = \frac{0.20}{0.60} = \frac{1}{3} \qquad \blacktriangle$$

Aunque presentamos la fórmula para $P(B|A)$ por medio de un ejemplo donde las posibilidades son igualmente probables, esto no constituye un requisito para su uso.

EJEMPLO 2.17

Con respecto a los dados amañados del ejemplo 2.9, ¿cuál es la probabilidad de que el número de puntos tirados sea un cuadrado perfecto? También, ¿cuál es la probabilidad de que sea un cuadrado perfecto dado que es mayor que 3?

Solución

Si A es el evento de que el número de puntos tirados sea mayor que 3 y B es el evento de que es un cuadrado perfecto, tenemos $A = \{4, 5, 6\}$, $B = \{1, 4\}$ y $A \cap B = \{4\}$. Puesto que las probabilidades de tirar un 1, 2, 3, 4, 5 o 6 con el dado son $\frac{2}{9}, \frac{1}{9}, \frac{2}{9}, \frac{1}{9}, \frac{2}{9}$ y $\frac{1}{9}$ (véase página 39), encontramos que la respuesta a la primera pregunta es

$$P(B) = \frac{2}{9} + \frac{1}{9} = \frac{1}{3}$$

Para determinar $P(B|A)$, calculamos primero

$$P(A \cap B) = \frac{1}{9} \quad \text{y} \quad P(A) = \frac{1}{9} + \frac{2}{9} + \frac{1}{9} = \frac{4}{9}$$

Entonces, al sustituir en la fórmula de la definición 2.1, obtenemos

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{\frac{1}{9}}{\frac{4}{9}} = \frac{1}{4} \qquad \blacktriangle$$

Para verificar que la fórmula de la definición 2.12 ha dado la respuesta "correcta" en el ejemplo precedente, sólo tenemos que asignar probabilidad v a los dos números pares en el espacio muestral reducido A y probabilidad $2v$ al número impar, de tal manera que la suma de las tres probabilidades sea igual a 1. Así tenemos $v + 2v + v = 1$, $v = \frac{1}{4}$ y, por tanto, $P(B|A) = \frac{1}{4}$ como antes.

EJEMPLO 2.18

Un fabricante de partes de aeroplano sabe, por experiencia, que la probabilidad de que una orden esté lista para embarque a tiempo es 0.80, y de que esté lista para embarque a tiempo y también se entregue a tiempo es 0.72. ¿Cuál es la probabilidad de que tal orden se entregue a tiempo dado que estuvo lista para embarque a tiempo?

Solución

Si R representa el evento de que una orden esté lista para embarque a tiempo y D sea el evento que se entregue a tiempo, tenemos $P(R) = 0.80$ y $P(R \cap D) = 0.72$, y resulta que

$$P(D|R) = \frac{P(R \cap D)}{P(R)} = \frac{0.72}{0.80} = 0.90$$

Así, 90% de los embarques se entregarán a tiempo con tal que sean embarcados a tiempo. Advierta que $P(R|D)$, la probabilidad de que un embarque que se entregó a tiempo también estuvo listo para embarque a tiempo, no se puede determinar sin información adicional; para este propósito también tendríamos que saber $P(D)$. ▲

Si multiplicamos las expresiones de ambos lados de la fórmula de la definición 2.1 por $P(A)$, obtenemos la siguiente **regla de multiplicación**.

> TEOREMA **2.9** Si A y B son dos eventos cualquiera en un espacio muestral S y $P(A) \neq 0$, entonces
>
> $$P(A \cap B) = P(A) \cdot P(B|A)$$

En palabras, la probabilidad de que A y B ocurran ambos es el producto de la probabilidad de A y la probabilidad condicional de B dado A. Alternativamente, si $P(B) \neq 0$, la probabilidad de que A y B ocurran ambos es el producto de la probabilidad de B y la probabilidad condicional de A dado B; con símbolos

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

Para derivar esta regla alternativa de multiplicación, intercambiamos A y B en la fórmula del teorema 2.9 y nos valemos del hecho que $A \cap B = B \cap A$.

EJEMPLO 2.19

Si seleccionamos aleatoriamente dos cinescopios en sucesión de un embarque de 240 cinescopios, de los cuales 15 están defectuosos, ¿cuál es la probabilidad de que ambos estén defectuosos?

Solución

Si suponemos probabilidades iguales para cada selección (que es lo que queremos decir al seleccionar aleatoriamente cinescopios), la probabilidad de que el primer cinescopio esté defectuoso es $\frac{15}{240}$, y la probabilidad de que el segundo cinescopio esté defectuoso dado que el primer cinescopio está defectuoso es $\frac{14}{239}$. Así la probabilidad de que ambos cinescopios estén defectuosos es $\frac{15}{240} \cdot \frac{14}{239} = \frac{7}{1{,}912}$.

Esto supone que estamos **muestreando sin reemplazo**; esto es, el primer cinescopio no se reemplaza antes de que se seleccione el segundo cinescopio. ▲

EJEMPLO 2.20

Encuentre las probabilidades de sacar aleatoriamente dos ases de una baraja ordinaria de 52 cartas de juego si muestreamos

(a) sin reemplazo;

(b) con reemplazo.

Solución

(a) Si la primera carta no se reemplaza antes de que se saque la segunda, la probabilidad de sacar dos ases en sucesión es

$$\frac{4}{52} \cdot \frac{3}{51} = \frac{1}{221}$$

(b) Si la primera carta se reemplaza antes de que se saque la segunda, la probabilidad correspondiente es

$$\frac{4}{52} \cdot \frac{4}{52} = \frac{1}{169} \qquad \blacktriangle$$

En las situaciones descritas en los dos ejemplos precedentes hay un orden temporal definido entre los dos eventos A y B. En general, éste no tiene que ser el caso cuando escribimos $P(A|B)$ o $P(B|A)$. Por ejemplo, podríamos preguntar por la probabilidad de que la primera carta que se sacó sea un as dado que la segunda carta que se sacó (sin reemplazo) es un as: la respuesta también sería $\frac{3}{51}$.

El teorema 2.9 se puede generalizar fácilmente para que sea válido en más de dos eventos; por ejemplo, para tres eventos tenemos

TEOREMA 2.10 Si A, B y C son tres eventos cualquiera en un espacio muestral S tal que $P(A \cap B) \neq 0$, entonces

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

Demostración. Escribiendo $A \cap B \cap C$ como $(A \cap B) \cap C$ y usando dos veces la fórmula del teorema 2.9, obtenemos

$$\begin{aligned} P(A \cap B \cap C) &= P[(A \cap B) \cap C] \\ &= P(A \cap B) \cdot P(C|A \cap B) \\ &= P(A) \cdot P(B|A) \cdot P(C|A \cap B) \qquad \blacktriangledown \end{aligned}$$

EJEMPLO 2.21

Una caja de fusibles contiene 20 fusibles, de los cuales cinco están defectuosos. Si se seleccionan tres fusibles aleatoriamente y se sacan de la caja en sucesión sin reemplazo, ¿cuál es la probabilidad de que los tres fusibles estén defectuosos?

Solución

Si A es el evento de que el primer fusible esté defectuoso, B es el evento de que el segundo fusible esté defectuoso, y C es el evento de que el tercer fusible esté defectuoso, entonces $P(A) = \frac{5}{20}$, $P(B|A) = \frac{4}{19}$, $P(C|A \cap B) = \frac{3}{18}$ y la sustitución en la fórmula nos da

$$P(A \cap B \cap C) = \frac{5}{20} \cdot \frac{4}{19} \cdot \frac{3}{18}$$
$$= \frac{1}{114} \quad \blacktriangle$$

Generalizaciones adicionales de los teoremas 2.9 y 2.10 a k eventos son simples, y la fórmula resultante se puede demostrar por inducción matemática.

2.7 EVENTOS INDEPENDIENTES

Hablando de manera informal, dos eventos A y B son **independientes** si la ocurrencia o no ocurrencia de cualquiera de los dos no afecta la probabilidad de la ocurrencia del otro. Para ilustrarlo tomemos el ejemplo precedente, donde las selecciones habrían sido independientes si cada fusible se hubiera reemplazado antes de que el siguiente se seleccionara; la probabilidad de sacar un fusible defectuoso habría permanecido $\frac{5}{20}$.

Con símbolos, dos eventos A y B son independientes si $P(B|A) = P(B)$ y $P(A|B) = P(A)$, y se puede demostrar que cualquiera de estas igualdades implica a la otra cuando ambas probabilidades condicionales existen, esto es, cuando ni $P(A)$ ni $P(B)$ es igual a cero (véase el ejercicio 2.65).

Ahora, si sustituimos $P(B)$ por $P(B|A)$ en la fórmula del teorema 2.9, obtenemos

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

y usaremos esto como nuestra definición formal de independencia.

DEFINICIÓN 2.2 Dos eventos A y B son **independientes** si y sólo si

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

Siguiendo los pasos en sentido inverso, podemos demostrar que la definición 2.2 implica la definición de independencia que dimos anteriormente.

Si dos eventos no son independientes, decimos que son **dependientes**. Al obtener la fórmula de la definición 2.2, suponemos que $P(B|A)$ existe y, por tanto, que $P(A) \neq 0$. Por conveniencia matemática, permitiremos que la definición también sea válida cuando $P(A) = 0$ y/o $P(B) = 0$.

EJEMPLO 2.22

Se lanza una moneda tres veces y se supone que los ocho resultados posibles, HHH, HHT, HTH, THH, HTT, THT, TTH y TTT, son igualmente probables. Si A es el evento de que una cara ocurra en cada uno de los dos primeros lanzamientos, B es el evento que una cruz ocurra en el tercer lanzamiento y C es el evento que exactamente dos cruces ocurren en los tres lanzamientos, demuestre que

(a) los eventos A y B son independientes;
(b) los eventos B y C son dependientes.

Solución

Puesto que

$$\begin{aligned} A &= \{\text{HHH, HHT}\} \\ B &= \{\text{HHT, HTT, THT, TTT}\} \\ C &= \{\text{HTT, THT, TTH}\} \\ A \cap B &= \{\text{HHT}\} \\ B \cap C &= \{\text{HTT, THT}\} \end{aligned}$$

el supuesto de que los ocho resultados posibles son todos equiprobables nos da $P(A) = \frac{1}{4}$, $P(B) = \frac{1}{2}$, $P(C) = \frac{3}{8}$, $P(A \cap B) = \frac{1}{8}$ y $P(B \cap C) = \frac{1}{4}$.

(a) Puesto que $P(A) \cdot P(B) = \frac{1}{4} \cdot \frac{1}{2} = \frac{1}{8} = P(A \cap B)$, los eventos A y B son independientes.
(b) Puesto que $P(B) \cdot P(C) = \frac{1}{2} \cdot \frac{3}{8} = \frac{3}{16} \neq P(B \cap C)$, los eventos A y B no son independientes. ▲

En relación con la definición 2.2, se puede demostrar que si A y B son independientes, entonces también lo son A y B', A' y B, y A' y B'. Por ejemplo,

TEOREMA 2.11 Si A y B son independientes, entonces A y B' también son independientes.

Demostración. Puesto que $A = (A \cap B) \cup (A \cap B')$, como se le pidió al lector que demostrara en el inciso (a) del ejercicio 2.4, $A \cap B$ y $A \cap B'$ son mutuamente excluyentes, y A y B son independientes por suposición, tenemos

$$\begin{aligned} P(A) &= P[(A \cap B) \cup (A \cap B')] \\ &= P(A \cap B) + P(A \cap B') \\ &= P(A) \cdot P(B) + P(A \cap B') \end{aligned}$$

Resulta que

$$\begin{aligned} P(A \cap B') &= P(A) - P(A) \cdot P(B) \\ &= P(A) \cdot [1 - P(B)] \\ &= P(A) \cdot P(B') \end{aligned}$$

y de ahí que A y B' sean independientes. ▼

En los ejercicios 2.66 y 2.67 se le pedirá al lector que demuestre que si A y B son independientes, entonces A' y B son independientes y así lo son A' y B', y si A y B son dependientes, entonces A y B' son dependientes.

Para extender el concepto de independencia a más de dos eventos, definamos lo siguiente

> **DEFINICIÓN 2.3** Los eventos $A_1, A_2, \ldots,$ y A_k son **independientes** si y sólo si la probabilidad de la intersección de cualesquiera 2, 3, ..., o k de estos eventos es igual al producto de sus probabilidades respectivas.

Para tres eventos A, B y C, por ejemplo, la independencia requiere que

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

$$P(A \cap C) = P(A) \cdot P(C)$$

$$P(B \cap C) = P(B) \cdot P(C)$$

y

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

Es de interés señalar que tres o más eventos pueden ser **independientes por parejas** sin ser independientes.

EJEMPLO 2.23

La figura 2.11 muestra un diagrama de Venn con probabilidades asignadas a sus diversas regiones. Verifique que A y B son independientes, que A y C son independientes y que B y C son independientes pero que A, B y C no son independientes.

Solución

Como se puede ver en el diagrama, $P(A) = P(B) = P(C) = \frac{1}{2}$, $P(A \cap B) = P(A \cap C) = P(B \cap C) = \frac{1}{4}$ y $P(A \cap B \cap C) = \frac{1}{4}$. Así,

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

$$P(A) \cdot P(C) = \frac{1}{4} = P(A \cap C)$$

$$P(B) \cdot P(C) = \frac{1}{4} = P(B \cap C)$$

pero

$$P(A) \cdot P(B) \cdot P(C) = \frac{1}{8} \neq P(A \cap B \cap C) \qquad \blacktriangle$$

Figura 2.11 Diagrama de Venn para el ejemplo 2.23.

A propósito, al ejemplo anterior se le puede dar una interpretación "real" al considerar un cuarto grande que tiene tres interruptores separados que controlan las luces del techo. Estas luces estarán encendidas cuando los tres interruptores estén "hacia arriba" y por tanto también cuando uno de los interruptores esté "hacia arriba" y los otros dos estén "hacia abajo". Si A es el evento que el primer interruptor esté "hacia arriba", B es el evento que el segundo interruptor esté "hacia arriba", y C es el evento que el tercer interruptor esté "hacia arriba", el diagrama de Venn de la figura 2.11 muestra un posible conjunto de probabilidades asociado con que los interruptores estén "hacia arriba" o "hacia abajo" cuando las luces del techo están prendidas.

También puede suceder que $P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)$ sin que A, B y C sean independientes por parejas; se le pedirá al lector que verifique esto en el ejercicio 2.68.

Por supuesto, si se nos da que ciertos eventos son independientes, la probabilidad de que todos ocurran es simplemente el producto de sus respectivas probabilidades.

EJEMPLO 2.24

Encuentre las probabilidades de obtener

(a) tres caras en tres lanzamientos aleatorios de una moneda balanceada;

(b) cuatro seis y después otro número en cinco lanzamientos aleatorios de un dado balanceado.

Solución

(a) Al multiplicar las probabilidades respectivas, obtenemos

$$\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$$

(b) Al multiplicar las probabilidades respectivas, obtenemos

$$\frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{5}{6} = \frac{5}{7{,}776} \qquad \blacktriangle$$

2.8 TEOREMA DE BAYES

En muchas situaciones el resultado de un experimento depende de lo que sucede en varias etapas intermedias. Lo siguiente es un ejemplo sencillo donde hay una etapa intermedia que consta de dos alternativas:

EJEMPLO 2.25

La terminación de un trabajo de construcción se puede retrasar a causa de una huelga. Las probabilidades son 0.60 de que habrá una huelga, 0.85 de que el trabajo de construcción se termine a tiempo si no hay huelga y 0.35 que el trabajo de construcción se termine a tiempo si hay huelga. ¿Cuál es la probabilidad de que el trabajo de construcción se termine a tiempo?

Solución

Si A es el evento de que el trabajo de construcción se terminará a tiempo y B es el evento de que habrá una huelga, se nos dan $P(B) = 0.60$, $P(A|B') = 0.85$ y $P(A|B) = 0.35$. Nos valemos de la fórmula del inciso (a) del ejercicio 2.4, del hecho de que $A \cap B$ y $A \cap B'$ son mutuamente excluyentes y de la forma alternativa de la regla de multiplicación, podemos escribir

$$\begin{aligned} P(A) &= P[(A \cap B) \cup (A \cap B')] \\ &= P(A \cap B) + P(A \cap B') \\ &= P(B) \cdot P(A|B) + P(B') \cdot P(A|B') \end{aligned}$$

Entonces, al sustituir los valores numéricos dados, obtenemos

$$\begin{aligned} P(A) &= (0.60)(0.35) + (1 - 0.60)(0.85) \\ &= 0.55 \qquad \blacktriangle \end{aligned}$$

Una generalización inmediata de esta clase de situación es el caso donde la etapa intermedia permite k alternativas diferentes (cuya ocurrencia se denota por B_1,

$B_2, \ldots, B_k$). Requiere el siguiente teorema, algunas veces llamado la **regla de la probabilidad total** o la **regla de eliminación**.

TEOREMA 2.12 Si los eventos $B_1, B_2, \ldots,$ y B_k constituyen una partición del espacio muestral S y $P(B_i) \neq 0$ para $i = 1, 2, \ldots, k$, entonces para cualquier evento A en S

$$P(A) = \sum_{i=1}^{k} P(B_i) \cdot P(A|B_i)$$

Como fue definido en la nota al pie de la página 10, las B constituyen una partición del espacio muestral si son mutuamente excluyentes por parejas y si su unión es igual a S. Una demostración formal del teorema 2.12 consta, esencialmente, de los mismos pasos del ejemplo 2.25, y se le deja al lector en el ejercicio 2.74.

EJEMPLO 2.26

Los miembros de una empresa de consultoría rentan automóviles de tres agencias de renta de automóviles: 60% de la agencia 1, 30% de la agencia 2, y 10% de la agencia 3. Si 9% de los automóviles de la agencia 1 necesita una afinación, 20% de los autos de la agencia 2 necesitan una afinación, y 6% de los autos de la agencia 3 necesitan una afinación, ¿cuál es la probabilidad de que un automóvil rentado, entregado a la empresa, necesite una afinación?

Solución

Si A es el evento de que el automóvil necesita una afinación, y B_1, B_2 y B_3 son los eventos de que el automóvil venga de las agencias 1, 2 o 3, tenemos $P(B_1) = 0.60$, $P(B_2) = 0.30$, $P(B_3) = 0.10$, $P(A|B_1) = 0.09$, $P(A|B_2) = 0.20$ y $P(A|B_3) = 0.06$. Al sustituir estos valores en la fórmula del teorema 2.12 obtenemos

$$\begin{aligned} P(A) &= (0.60)(0.09) + (0.30)(0.20) + (0.10)(0.06) \\ &= 0.12 \end{aligned}$$

Así, 12% de todos los automóviles rentados entregados a esta empresa necesitarán una afinación. ▲

Con respecto al ejemplo precedente, supongamos que nos interesa la siguiente pregunta: si un automóvil rentado entregado a la empresa de consultoría necesita una afinación, ¿cuál es la probabilidad de que haya venido de la agencia de renta 2? Para responder a preguntas de esta clase, necesitamos el siguiente teorema, llamado el **teorema de Bayes**:

TEOREMA 2.13 Si $B_1, B_2, \ldots, B_k$ constituye una partición del espacio muestral S y $P(B_i) \neq 0$ para $i = 1, 2, \ldots, k$, entonces para cualquier evento A en S tal que $P(A) \neq 0$

$$P(B_r|A) = \frac{P(B_r) \cdot P(A|B_r)}{\sum_{i=1}^{k} P(B_i) \cdot P(A|B_i)}$$

para $r = 1, 2, \ldots, k$.

En palabras, la probabilidad de que el evento A se haya alcanzado a través de la $r$ésima rama del diagrama de árbol de la figura 2.12, dado que se alcanzó a través de una de sus k ramas, es la *razón* de la probabilidad asociada con la $r$ésima rama a la suma de las probabilidades asociadas con todas las k ramas del árbol.

Demostración. Escribimos $P(B_r|A) = \dfrac{P(A \cap B_r)}{P(A)}$ de acuerdo a la definición de probabilidad condicional, sólo tenemos que sustituir $P(B_r) \cdot P(A|B_r)$ con $P(A \cap B_r)$ y la fórmula del teorema 2.12 por $P(A)$. ▼

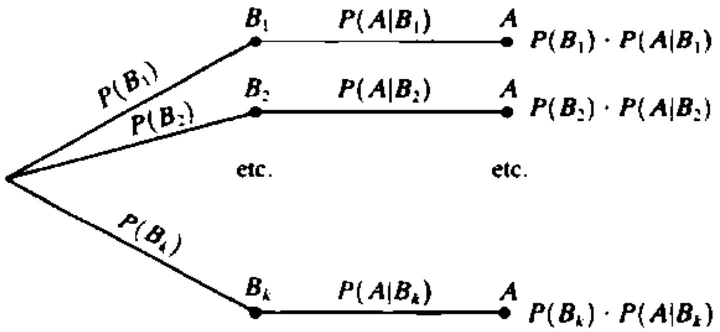

Figura 2.12 Diagrama de árbol para el teorema de Bayes.

EJEMPLO 2.27

Con respecto al ejemplo 2.26, si un automóvil de renta entregado a la empresa de consultoría necesita una afinación, ¿cuál es la probabilidad de que provenga de la agencia de renta 2?

Solución

Al sustituir las probabilidades en la página 63 en la fórmula del teorema 2.13, obtenemos

$$\begin{aligned} P(B_2|A) &= \frac{(0.30)(0.20)}{(0.60)(0.09) + (0.30)(0.20) + (0.10)(0.06)} \\ &= \frac{0.060}{0.120} \\ &= 0.5 \end{aligned}$$

Observe que aunque sólo 30% de los automóviles entregados a la empresa vienen de la agencia 2, 50% de los automóviles que requieren una afinación vienen de esa agencia. ▲

EJEMPLO 2.28

En un cierto estado, 25% de todos los automóviles emiten cantidades excesivas de contaminantes. Si la probabilidad es 0.99 de que un auto que emite cantidades excesivas de contaminantes fallará las pruebas de emisión vehicular del estado, y la probabilidad es 0.17 que un automóvil que no emite cantidades excesivas de contaminantes aun así fallará la prueba, ¿cuál es la probabilidad de que un auto que falla la prueba emita cantidades excesivas de contaminantes?

Solución

Si dibujamos esta situación como en la figura 2.13, encontramos que las probabilidades asociadas con las dos ramas del diagrama de árbol son = $(0.25)(0.99) = 0.2475$ y $(1 - 0.25)(0.17) = 0.1275$. Así, la probabilidad de que un auto que falla la prueba realmente emita cantidades excesivas de contaminantes es

$$\frac{0.2475}{0.2475 + 0.1275} = 0.66$$

Por supuesto, este resultado también se podía haber obtenido sin el diagrama al sustituir directamente en la fórmula del teorema de Bayes. ▲

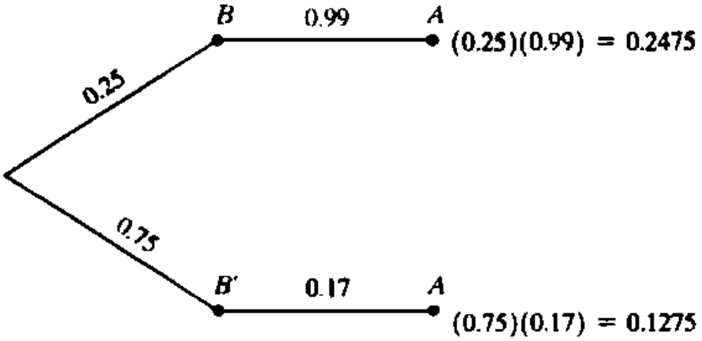

Figura 2.13 Diagrama de árbol para el ejemplo 2.28.

Aunque el teorema de Bayes resulta de los postulados de probabilidad y la definición de probabilidad condicional, ha sido objeto de una amplia controversia. Es indudable la validez del teorema de Bayes, pero numerosos argumentos han surgido sobre la asignación de **probabilidades previas** $P(B_i)$. También, una buena dosis de misticismo rodea al teorema de Bayes porque supone una forma de razonamiento "al revés", o a la "inversa", esto es, razonar "del efecto a la causa". Por ejemplo, fallar la prueba es el efecto y emitir cantidades excesivas de contaminantes es una causa posible (ejemplo 2.28).

EJERCICIOS

2.61 Demuestre que los postulados de probabilidad se satisfacen por las probabilidades condicionales. En otras palabras, demuestre que si $P(B) \neq 0$, entonces

(a) $P(A|B) \geqq 0$;

(b) $P(B|B) = 1$;

(c) $P(A_1 \cup A_2 \cup \cdots | B) = P(A_1|B) + P(A_2|B) + \cdots$ para cualquier secuencia de eventos mutuamente excluyentes $A_1, A_2, \ldots$.

2.62 Demuestre por medio de ejemplos numéricos que $P(B|A) + P(B|A')$
(a) puede ser igual a 1;
(b) no necesita ser igual a 1.

2.63 Repitiendo el método de demostración del teorema 2.10, muestre que

$$P(A \cap B \cap C \cap D) = P(A) \cdot P(B|A) \cdot P(C|A \cap B) \cdot P(D|A \cap B \cap C)$$

con tal que $P(A \cap B \cap C) \neq 0$.

2.64 Dados tres eventos A, B y C tales que $P(A \cap B \cap C) \neq 0$ y $P(C|A \cap B) = P(C|B)$, demuestre que $P(A|B \cap C) = P(A|B)$.

2.65 Demuestre que si $P(B|A) = P(B)$ y $P(B) \neq 0$, entonces $P(A|B) = P(A)$.

2.66 Demuestre que si los eventos A y B son independientes, entonces
(a) los eventos A' y B son independientes;
(b) los eventos A' y B' son independientes.

2.67 Demuestre que si los eventos A y B son dependientes, entonces los eventos A y B' son dependientes.

2.68 Tome la figura 2.14 como referencia para demostrar que $P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)$ no implica necesariamente que A, B, y C son todos independientes por parejas.

Figura 2.14 Diagrama para los ejercicios 2.68, 2.69 y 2.70.

2.69 Tome la figura 2.14 como referencia para demostrar que si A es independiente de B y A es independiente de C, entonces B no es necesariamente independiente de C.

2.70 Tome la figura 2.14 como referencia para demostrar que si A es independiente de B y A es independiente de C, entonces A no es necesariamente independiente de $B \cup C$.

2.71 Si los eventos A, B, y C son independientes, demuestre que
(a) A y $B \cap C$ son independientes;
(b) A y $B \cup C$ son independientes;

2.72 Demuestre que se deben satisfacer $2^k - k - 1$ condiciones para que k eventos sean independientes.

2.73 Para cualquier evento A, demuestre que A y $\emptyset$ son independientes.

2.74 Demuestre el teorema 2.12 valiéndose de la siguiente generalización de la ley distributiva dada en el inciso (b) del ejercicio 2.1:

$$A \cap (B_1 \cup B_2 \cup \cdots \cup B_k) = (A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_k)$$

APLICACIONES

2.75 Hay noventa aspirantes para un trabajo en el departamento de noticias de una estación de televisión. Algunos son egresados de la universidad y algunos no, algunos de ellos tienen al menos tres años de experiencia y algunos no la tienen, el análisis exacto es

	Egresados de la universidad	*No egresados de la universidad*
Al menos tres años de experiencia	18	9
Menos de tres años de experiencia	36	27

Si el orden en que el gerente de la estación entrevista a los aspirantes es aleatorio, G es el evento que el primer aspirante entrevistado sea un egresado de la universidad, y T es el evento de que el primer aspirante entrevistado tenga al menos tres años de experiencia, determine cada una de las siguientes probabilidades directamente de los asientos y de los renglones y columnas de la tabla:

(a) $P(G)$; (b) $P(T')$; (c) $P(G \cap T)$;
(d) $P(G' \cap T')$; (e) $P(T|G)$; (f) $P(G'|T')$.

2.76 Use los resultados de ejercicio 2.75 para verificar que

(a) $P(T|G) = \dfrac{P(G \cap T)}{P(G)}$;

(b) $P(G'|T') = \dfrac{P(G' \cap T')}{P(T')}$.

2.77 Con respecto al ejercicio 2.46, ¿cuál es la probabilidad de que el doctor escogido para representar al personal del hospital en la convención tenga seguro contra tratamiento erróneo dado que es un cirujano?

2.78 Con respecto al ejercicio 2.49, ¿cuál es la probabilidad de que un marido vote en la elección dado que su mujer va a votar?

2.79 Con respecto al ejercicio 2.51, ¿cuál es la probabilidad de que uno de los estudiantes viva en el campus dado que es de fuera del estado?

2.80 Una caja contiene 100 pelotas, de las cuales 25 son rojas, 40 son blancas, y 35 son negras. Si se seleccionan dos pelotas de la caja, ¿cuál es la probabilidad que una será roja y una será blanca

(a) si la primera pelota se reemplaza antes de sacar la segunda;

(b) si la segunda pelota se saca sin reemplazar la primera?

2.81 Se cree que las probabilidades son 0.20, 0.40, 0.30 y 0.10 de que los equipos de baloncesto de cuatro universidades, T, U, V y W, ganen el campeonato de su conferencia. Si se pone a la universidad U bajo vigilancia y se le declara inelegible para el campeonato, ¿cuál es la probabilidad de que la universidad T gane el campeonato de la conferencia?

2.82 Con respecto al ejercicio 2.52, encuentre las probabilidades de que una persona que visite Disneylandia

(a) viajará en el Monorriel dado que irá en Crucero de la Jungla;

(b) tomará el paseo del Matterhorn dado que irá al Crucero de la Jungla y viajará en el Monorriel;

(c) no irá al Crucero de la Jungla dado que viajará en el Monorriel y/o tomará el paseo del Matterhorn;

(d) tomará el paseo del Matterhorn e irá al Crucero de la Jungla dado que no viajará en el Monorriel.

(*Sugerencia*: Dibuje un diagrama de Venn y anote las probabilidades asociadas con las diversas regiones.)

2.83 La probabilidad de sobrevivir a una cierta operación de trasplante es 0.55. Si un paciente sobrevive la operación, la probabilidad de que su cuerpo rechace el trasplante en menos de un mes es 0.20. ¿Cuál es la probabilidad de que sobreviva a estas etapas críticas?

2.84 Se examinan canastas o cajas de huevo en busca de coágulos de sangre al remover aleatoriamente tres huevos sucesivamente y examinar su contenido. Si los tres huevos están bien, se embarca la caja; de otra manera se rechaza. ¿Cuál es la probabilidad de que una caja se embarque si contiene 120 huevos, de los cuales 10 tienen coágulos de sangre?

2.85 Supongamos que en Vancouver, B. C., la probabilidad de que a un día lluvioso de otoño le siga un día lluvioso es 0.80 y la probabilidad de que a un día soleado le siga un día lluvioso es 0.60. Encontrar las probabilidades de que a un día lluvioso de otoño le siga

(a) un día lluvioso, un día soleado y otro día lluvioso;

(b) dos días soleados y después un día lluvioso;

(c) dos días lluviosos y después dos días soleados;

(d) llueva dos días mas tarde.

[*Sugerencia*: En el inciso (c) use la fórmula del ejercicio 2.63.]

2.86 Use la fórmula del ejercicio 2.63 para encontrar la probabilidad de escoger aleatoriamente (sin reemplazo) cuatro conejillos de Indias de una jaula que contiene 20 conejillos de Indias, de los cuales 15 están sanos y 5 están enfermos.

2.87 Se lanza dos veces un dado balanceado. Si A es el evento de que en el primer tiro salga un número par, B es el evento de que en el segundo tiro salga un número par, y C es el evento de que ambos tiros resulten en el mismo número, ¿son los eventos A, B y C

(a) independientes por parejas,

(b) independientes?

2.88 Un tirador certero acierta en el blanco con una probabilidad de 0.75. Si se supone independencia, encuentre las probabilidades de obtener

(a) un acierto seguido de dos fallas;

(b) dos aciertos y una falla en cualquier orden.

2.89 Una moneda está arreglada de manera que las probabilidades de cara y cruz son 0.52 y 0.48 respectivamente. Si la moneda se lanza tres veces, ¿cuáles son las probabilidades de sacar

(a) sólo caras;

(b) dos cruces y una cara en ese orden?

2.90 Un embarque de 1,000 partes contiene 1% de partes defectuosas. Encuentre la probabilidad de que

(a) las primeras cuatro partes escogidas arbitrariamente para inspección no sean defectuosas;

(b) la primera parte defectuosa se encontrará en la cuarta inspección.

2.91 Los registros médicos muestran que una entre 10 personas en una cierta ciudad tiene deficiencia tiroidea. Si se escogen aleatoriamente 12 personas en esta ciudad y se les hacen análisis, ¿cuál es la probabilidad de que al menos una de ellas tenga una deficiencia tiroidea?

2.92 Si cinco de los 10 camiones repartidores de una compañía no satisfacen los estándares de emisión y tres de ellos se seleccionan para una inspección, ¿cuál es la probabilidad de que ninguno de los camiones seleccionados satisfará los estándares de emisión?

2.93 Si una persona escoge aleatoriamente cuatro de 15 monedas de oro que un cambista tiene en almacén, y seis de las monedas son falsas, ¿cuál es la probabilidad de que las monedas escogidas sean todas falsas?

2.94 Una tienda departamental que factura a sus clientes una vez al mes ha encontrado que si un cliente paga oportunamente en un mes, la probabilidad es 0.90 de que él o ella también pague oportunamente el siguiente mes; sin embargo, si un cliente no paga oportunamente en un mes, la probabilidad de que él o ella pague oportunamente el siguiente mes es solamente 0.40.

(a) ¿Cuál es la probabilidad de que un cliente que paga oportunamente en un mes también pagará oportunamente los tres meses siguientes?

(b) ¿Cuál es la probabilidad de que un cliente que no paga oportunamente en un mes tampoco pagará oportunamente los siguientes dos meses y después haga un pago oportuno al mes siguiente de ello?

2.95 Con respecto a la figura 2.15, verifique que los eventos A, B, C y D son independientes. Advierta que la región que representa a A consta de dos círculos, y también es así para las regiones que representan B y C.

2.96 En una planta electrónica, se sabe por experiencia previa que la probabilidad es 0.84 de que un trabajador nuevo que haya asistido al programa de capacita-

Figura 2.15 Diagrama para el ejercicio 2.95.

ción o adiestramiento de la compañía cumplirá con su cuota de producción, y la probabilidad correspondiente es 0.49 para un trabajador nuevo que no haya asistido al programa de capacitación de la compañía. Si 70% de todos los trabajadores nuevos asisten al programa de capacitación, ¿cuál es la probabilidad de que un trabajador nuevo cumplirá con la cuota de producción?

2.97 En un laberinto T, a una rata se le da comida si voltea a la izquierda y una descarga de electricidad si voltea a la derecha. En el primer ensayo hay una probabilidad de 50-50 de que la rata se voltee en cualquier dirección; entonces, si recibe alimento en el primer ensayo, la probabilidad es 0.68 de que volteará a la izquierda en el siguiente ensayo, y si recibe una descarga eléctrica en el primer ensayo, la probabilidad es 0.84 de que volteará a la izquierda en el siguiente ensayo. ¿Cuál es la probabilidad de que la rata voltee a la izquierda en el segundo ensayo?

2.98 Por experiencia se sabe que en una cierta industria 60% de todos los litigios entre los trabajadores y la administración son por salarios, 15% por las condiciones de trabajo y 25% son sobre aspectos de prestaciones. También 45% de los litigios por salarios se resuelven sin huelgas, 70% de los litigios por condiciones de trabajo se resuelven sin huelgas y 40% de los litigios acerca de prestaciones se resuelven sin huelgas. ¿Cuál es la probabilidad de que un litigio entre trabajadores y la administración se resuelva sin una huelga?

2.99 Con respecto al ejercicio 2.98, ¿cuál es la probabilidad de que si un litigio entre los trabajadores y la administración se resuelva sin una huelga, sea por salarios?

2.100 La probabilidad de que un accidente de un solo automóvil sea debido a frenos defectuosos es 0.04, la probabilidad de que un accidente de un solo auto sea correctamente atribuido a frenos defectuosos es 0.82, y la probabilidad de que un accidente de un solo auto sea incorrectamente atribuido a frenos defectuosos es 0.03. ¿Cuál es la probabilidad de que

(a) un accidente de un solo auto será atribuido a frenos defectuosos;

(b) un accidente de un solo auto atribuido a frenos defectuosos fue realmente causado por frenos defectuosos?

2.101 En una cierta comunidad, 8% de todos los adultos mayores de 50 años tienen diabetes. Si un servicio de salud en esta comunidad diagnostica correctamente a 95% de las personas con diabetes como enfermas de diabetes e incorrectamente diagnostica a 2% de todas las personas sin diabetes como enfermas de diabetes, encuentre las probabilidades de que

(a) el servicio de salud comunitario diagnosticará a un adulto mayor de 50 como enfermo de diabetes;

(b) una persona mayor de 50 diagnosticada con diabetes por el servicio de salud realmente tenga la enfermedad.

2.102 Con respecto al ejemplo 2.25, suponga que descubrimos posteriormente que el trabajo se terminó a tiempo. ¿Cuál es la probabilidad de que haya habido una huelga?

2.103 Una casa de ventas por correo emplea tres dependientes de almacén, U, V y W, quienes retiran artículos de los anaqueles y los reúnen para su subsecuente verificación y empaque. U comete un error en una orden (saca el artículo equivocado o la cantidad equivocada) una vez de cien, V comete un error en una orden cinco veces de cien, y W comete un error en una orden tres veces de cien. Si U, V y W surten, respectivamente, 30, 40 y 30% de todas las órdenes, ¿cuál es la probabilidad de que

(a) se cometa un error en una orden;

(b) si se comete un error en una orden, la orden haya sido surtida por U;

(c) si se comete un error en una orden, la orden haya sido surtida por V?

2.104 La explosión en una obra en construcción pudo haber ocurrido como resultado de electricidad estática, mal funcionamiento del equipo, descuido o sabotaje. Entrevistas con los ingenieros de construcción que analizaron los riesgos implicados dieron como resultado cálculos de que tal explosión ocurriera con probabilidad de 0.25 como un resultado de la electricidad estática, 0.20 como un resultado del mal funcionamiento del equipo, 0.40 como un resultado de descuido, y 0.75 como un resultado de sabotaje. También se juzgó que las probabilidades previas de las cuatro causas de la explosión eran 0.20, 0.40, 0.25 y 0.15. Con base en toda esta información, ¿cuál es

(a) la causa más verosímil de la explosión;

(b) la causa menos verosímil de la explosión?

2.105 Una corredora de arte recibe un embarque de cinco pinturas antiguas del extranjero, y sobre la base de la experiencia pasada estima que las probabilidades son, respectivamente, 0.76, 0.09, 0.02, 0.01, 0.02 y 0.10 de que 0, 1, 2, 3, 4 o todas las 5 sean falsificaciones. Puesto que el costo de autenticación es bastante alto, decide seleccionar aleatoriamente una de las cinco pinturas y enviarla para su autenticación. Si resulta que esta pintura es una falsificación, ¿qué probabilidad debe asignarle ahora a la posibilidad de que todas las otras pinturas sean también falsificaciones?

2.106 Para obtener respuestas a preguntas delicadas, algunas veces usamos un método llamado la **técnica de respuesta al azar**. Supongamos, por ejemplo, que queremos determinar qué porcentaje de los estudiantes a una universidad grande fuman marihuana. Elaboramos 20 tarjetas "flash", anotamos "Fumo mariguana al menos una vez por semana" en 12 de las tarjetas, donde 12 se selecciona arbitrariamente, y "No fumo mariguana al menos una vez por semana" en las otras. Entonces, dejamos que cada estudiante (en la muestra entrevistada) seleccione al azar una de las 20 tarjetas y responda "sí" o "no" sin divulgar la pregunta.

(a) Establezca una relación entre $P(Y)$, la probabilidad de que un estudiante dará una respuesta de "sí" y $P(M)$, la probabilidad de que un estudiante seleccionado aleatoriamente en esa universidad fume marihuana al menos una vez por semana.

(b) Si 106 de 250 estudiantes respondieron "sí" bajo estas condiciones, use el resultado del inciso (a) y $\frac{106}{250}$ como un estimado de $P(Y)$ para estimar $P(M)$.

REFERENCIAS

Entre los numerosos libros de texto sobre teoría de la probabilidad publicados en años recientes, uno de los más populares es

FELLER, W., *An Introduction to Probability Theory and Its Applications*, Vol. I, 3rd ed. Nueva York: John Wiley & Sons, Inc., 1968.

Se pueden encontrar tratamientos más elementales en

BARR, D. R., and ZEHNA, P. W., *Probability: Modeling Uncertainty*. Reading, Mass.: Addison-Wesley Publishing Company, Inc., 1983,

DRAPER, N. R., and LAWRENCE, W. E., *Probability: An Introductory Course*. Chicago: Markam Publishing Company, 1970,

FREUND, J. E., *Introduction to Probability*. Encino, Calif.: Dickenson Publishing Co., Inc., 1973,

GOLDBERG, S., *Probability—An Introduction*. Mineola, N.Y.: Dover Publications, Inc. (Reedición de la edición de 1960),

HODGES, J. L., and LEHMANN, E. L., *Elements of Finite Probability*. San Francisco: Holden-Day, Inc., 1965,

NOSAL, M., *Basic Probability and Applications*. Philadelphia: W. B. Saunders Company, 1977.

En muchos textos se dan tratamientos más avanzados, por ejemplo,

HOEL, P., PORT, S. C., and STONE, C. J., *Introduction to Probability Theory*. Boston: Houghton Mifflin Company, 1971,

KHAZANIE, R., *Basic Probability Theory and Applications*. Pacific Palisades, Calif.: Goodyear Publishing Company, Inc., 1976,

PARZEN, E., *Modern Probability Theory and Its Applications*. Nueva York: John Wiley & Sons, Inc., 1960,

ROSS, S., *A First Course in Probability*, 3rd ed. Nueva York: Macmillan Publishing Company, 1988,

SOLOMON, F., *Probability and Stochastic Processes*. Upper Saddle River, N.J.: Prentice Hall, 1987.

CAPÍTULO

3

Distribuciones de probabilidad y densidades de probabilidad

3.1 INTRODUCCIÓN

En la mayoría de los problemas aplicados que contienen probabilidades sólo nos interesa un aspecto particular (o dos o unos cuantos aspectos particulares) de los resultados de los experimentos. Por ejemplo, cuando lanzamos un par de dados generalmente nos interesa sólo el total y no el resultado de cada dado: cuando entrevistamos a una pareja de casados seleccionada aleatoriamente quizá nos interese el tamaño de su familia y su ingreso combinado, pero no el número de años que han estado casados o el total de sus activos; y cuando muestreamos focos incandescentes producidos en serie tal vez nos interese su durabilidad o su brillantez, pero no su precio.

En cada uno de estos ejemplos nos interesan números asociados con los resultados de experimentos de azar, esto es, en los valores asumidos por las **variables aleatorias**. En el lenguaje de la probabilidad y la estadística, el total que lanzamos con un par de dados es una variable aleatoria, el tamaño de la familia de una pareja de casados escogida aleatoriamente y su ingreso combinado son variables aleatorias, y también lo son la durabilidad y la brillantez de un foco incandescente escogido aleatoriamente para inspección.

Para ser más explícitos, considere la figura 3.1, que (como la figura 2.1 en la página 29) representa el espacio muestral para un experimento en el que lanzamos un par de dados, y supongamos que cada uno de los 36 resultados posibles tiene la probabilidad $\frac{1}{36}$. Sin embargo, advierta que en la figura 3.1 hemos asociado un número a cada punto: por ejemplo, hemos asociado el número 2 al punto (1, 1), el número 6 al punto (1, 5), el número 8 al punto(6, 2), el número 11 al punto (5, 6) y así sucesivamente. Evi-

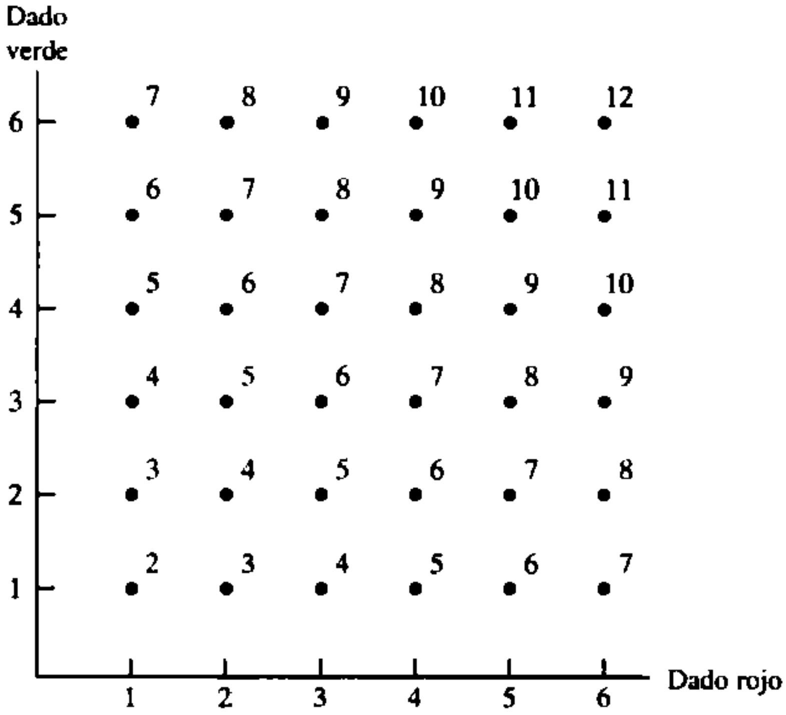

Figura 3.1 Número total de puntos obtenidos con un par de dados.

dentemente, asociamos con cada punto el valor de una variable aleatoria, esto es, el total correspondiente al lanzamiento del par de dados.

Puesto que "asociar un número con cada punto (elemento) de un espacio muestral" es simplemente otra forma de decir que estamos "definiendo una función sobre los puntos de un espacio muestral", hagamos ahora la siguiente definición.

> **DEFINICIÓN 3.1** Si S es un espacio muestral con una medida de probabilidad y X es una función de valor real definida sobre los elementos de S, entonces X se llama una **variable aleatoria**.†

En este libro siempre denotaremos las variables aleatorias con letras mayúsculas y sus valores con las correspondientes letras minúsculas; por ejemplo, escribiremos x para denotar el *valor* de la variable aleatoria X.

Con relación al ejemplo precedente y a la figura 3.1, observe que la variable aleatoria X asume el valor 9, y escribimos $X = 9$ para el subconjunto

$$\{(6,3), (5,4), (4,5), (3,6)\}$$

† En algunos libros se usan los términos "variable de probabilidad", "variable estocástica" y "variate" en vez de "variable aleatoria".

del espacio muestral S. Así, $X = 9$ se interpreta como el conjunto de elementos de S para el cual el total es 9, y, más generalmente, $X = x$ se debe interpretar como el conjunto de elementos del espacio muestral para el cual la variable aleatoria X asume el valor x. Esto puede parecer confuso, pero recuerda a uno de los matemáticos que dicen "$f(x)$ es una función de x".

EJEMPLO 3.1

Se seleccionan dos calcetines al azar y se sacan en sucesión de una gaveta que contiene cinco calcetines cafés y tres calcetines verdes. Enumere los elementos del espacio muestral, las probabilidades correspondientes y los valores w que corresponden a la variable aleatoria W, donde W es el número de calcetines cafés seleccionados.

Solución

Si B y G representan café y verde, las probabilidades para BB, BG, GB y GG son, respectivamente $\frac{5}{8}\cdot\frac{4}{7} = \frac{5}{14}$, $\frac{5}{8}\cdot\frac{3}{7} = \frac{15}{56}$, $\frac{3}{8}\cdot\frac{5}{7} = \frac{15}{56}$ y $\frac{3}{8}\cdot\frac{2}{7} = \frac{3}{28}$, y los resultados se muestran en la siguiente tabla

Elemento del espacio muestral	*Probabilidad*	w
BB	$\frac{5}{14}$	2
BG	$\frac{15}{56}$	1
GB	$\frac{15}{56}$	1
GG	$\frac{3}{28}$	0

También, podemos escribir $P(W = 2) = \frac{5}{14}$, por ejemplo, para la probabilidad del evento de que la variable aleatoria W asumirá el valor 2. ▲

EJEMPLO 3.2

Se lanza cuatro veces una moneda balanceada. Enumere los elementos del espacio muestral que se presume son igualmente probables, ya que esto es lo que queremos decir por una moneda que está balanceada, y los valores correspondientes de x de la variable aleatoria X, el número total de caras.

Solución

Si H y T representan cara y cruz, los resultados se muestran en la siguiente tabla:

Elemento del espacio muestral	*Probabilidad*	x
HHHH	$\frac{1}{16}$	4
HHHT	$\frac{1}{16}$	3
HHTH	$\frac{1}{16}$	3
HTHH	$\frac{1}{16}$	3
THHH	$\frac{1}{16}$	3
HHTT	$\frac{1}{16}$	2
HTHT	$\frac{1}{16}$	2
HTTH	$\frac{1}{16}$	2
THHT	$\frac{1}{16}$	2
THTH	$\frac{1}{16}$	2
TTHH	$\frac{1}{16}$	2
HTTT	$\frac{1}{16}$	1
THTT	$\frac{1}{16}$	1
TTHT	$\frac{1}{16}$	1
TTTH	$\frac{1}{16}$	1
TTTT	$\frac{1}{16}$	0

Así, podemos escribir $P(X = 3) = \frac{4}{16}$, por ejemplo, para la probabilidad del evento de que la variable aleatoria X asumirá el valor 3. ▲

El hecho que la definición 3.1 se limita a funciones de valores reales no impone restricción alguna. Si los números que queremos asignar a los resultados de un experimento son números complejos, siempre podemos considerar por separado las partes reales e imaginarias como valores asumidos por dos variables aleatorias. También, si queremos describir cuantitativamente los resultados de un experimento, digamos, si damos el color del cabello de una persona, podemos hacer arbitrariamente que las descripciones tengan valores reales al codificar los diversos colores, y quizá al representarlos con los números 1, 2, 3 y así sucesivamente.

En todos los ejemplos de esta sección hemos limitado nuestro análisis a espacios muestrales discretos, y por tanto a **variables aleatorias discretas**, a saber, variables aleatorias cuyo intervalo es finito o infinito numerable. Las variables aleatorias continuas definidas sobre espacios muestrales continuos se estudiarán en la sección 3.3.

3.2 DISTRIBUCIONES DE PROBABILIDADES

Como ya vimos en los ejemplos 3.1 y 3.2, la medida de probabilidad definida en un espacio muestral discreto automáticamente provee las probabilidades de que una variable aleatoria asumirá cualquier valor dado dentro de su intervalo.

Por ejemplo, una vez asignada la probabilidad $\frac{1}{36}$ a cada elemento del espacio muestral de la figura 3.1, inmediatamente encontramos que la variable aleatoria X, el total lanzado con el par de dados, asume el valor 9 con probabilidad $\frac{4}{36}$; como se describe en la página 74, $X = 9$ contiene cuatro elementos igualmente probables del espacio muestral. Las probabilidades asociadas con todos los valores posibles de X se muestran en la siguiente tabla:

x	$P(X = x)$
2	$\frac{1}{36}$
3	$\frac{2}{36}$
4	$\frac{3}{36}$
5	$\frac{4}{36}$
6	$\frac{5}{36}$
7	$\frac{6}{36}$
8	$\frac{5}{36}$
9	$\frac{4}{36}$
10	$\frac{3}{36}$
11	$\frac{2}{36}$
12	$\frac{1}{36}$

En vez de mostrar las probabilidades asociadas con los valores de una variable aleatoria en una tabla, como hicimos en la ilustración precedente, es generalmente preferible dar una fórmula, esto es, expresar las probabilidades por medio de una función tal que sus valores, $f(x)$, sean iguales a $P(X = x)$ para cada x dentro del rango de la variable aleatoria X. Por ejemplo, para el total obtenido con un par de dados escribiríamos

$$f(x) = \frac{6 - |x - 7|}{36} \qquad \text{para } x = 2, 3, \ldots, 12$$

como se puede verificar fácilmente por sustitución. Claramente

$$f(2) = \frac{6 - |2 - 7|}{36} = \frac{6 - 5}{36} = \frac{1}{36}$$

$$f(3) = \frac{6 - |3 - 7|}{36} = \frac{6 - 4}{36} = \frac{2}{36}$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

$$f(12) = \frac{6 - |12 - 7|}{36} = \frac{6 - 5}{36} = \frac{1}{36}$$

y todos estos valores concuerdan con los mostrados en la tabla precedente.

DEFINICIÓN 3.2 Si X es una variable aleatoria discreta, la función dada por $f(x) = P(X = x)$ para cada x dentro del intervalo de X se llama la **distribución de probabilidad** de X.

Basado en los postulados de probabilidad, se deduce inmediatamente que

TEOREMA 3.1 Una función puede servir como la distribución de probabilidad de una variable aleatoria discreta X si y sólo si sus valores $f(x)$, satisfacen las condiciones

1. $f(x) \geqq 0$ para cada valor dentro de su dominio;
2. $\sum_x f(x) = 1$, donde la suma se extiende sobre todos los valores dentro de su dominio.

EJEMPLO 3.3

Encuentre una fórmula para la distribución de probabilidad del número total de caras obtenidas en cuatro lanzamientos de una moneda balanceada.

Solución

Con base en las probabilidades en la tabla de la página 76, encontramos que $P(X = 0) = \frac{1}{16}$, $P(X = 1) = \frac{4}{16}$, $P(X = 2) = \frac{6}{16}$, $P(X = 3) = \frac{4}{16}$ y $P(X = 4) = \frac{1}{16}$. Al observar que los numeradores de estas cinco fracciones 1, 4, 6, 4 y 1 son los coeficientes binomiales $\binom{4}{0}$, $\binom{4}{1}$, $\binom{4}{2}$, $\binom{4}{3}$ y $\binom{4}{4}$, encontramos que la fórmula para la distribución de probabilidad se puede escribir como

$$f(x) = \frac{\binom{4}{x}}{16} \qquad \text{para } x = 0, 1, 2, 3, 4$$

En la sección 5.4 se da una justificación teórica para esta fórmula y un tratamiento más general para n tiros de una moneda balanceada. ▲

EJEMPLO 3.4

Verifique si la función dada por

$$f(x) = \frac{x + 2}{25} \qquad \text{para } x = 1, 2, 3, 4, 5$$

puede servir como distribución de probabilidad de una variable aleatoria discreta.

Solución

Al sustituir los diferentes valores de x, obtenemos $f(1) = \frac{3}{25}$, $f(2) = \frac{4}{25}$, $f(3) = \frac{5}{25}$, $f(4) = \frac{6}{25}$ y $f(5) = \frac{7}{25}$. Puesto que todos estos valores son no negativos se satisface la primera condición del teorema 3.1, y puesto que

$$\begin{aligned} f(1) + f(2) + f(3) + f(4) + f(5) &= \frac{3}{25} + \frac{4}{25} + \frac{5}{25} + \frac{6}{25} + \frac{7}{25} \\ &= 1 \end{aligned}$$

se satisface la segunda condición del teorema 3.1. Así, la función dada puede servir como la distribución de probabilidad de una variable aleatoria que tenga el intervalo $\{1, 2, 3, 4, 5\}$. Por supuesto, si alguna variable aleatoria dada realmente tiene esta distribución de probabilidad es un asunto totalmente distinto. ▲

En algunos problemas es deseable presentar gráficamente las distribuciones de probabilidad y las dos clases de presentaciones gráficas usadas para este propósito se muestran en las figuras 3.2 y 3.3. La mostrada en la figura 3.2, llamada un **histograma de probabilidad**, o simplemente un **histograma**, representa la distribución de probabilidad del ejemplo 3.3. La altura de cada rectángulo es igual a la probabilidad de que X asuma el valor que corresponde al punto medio de su base. Al representar 0 con el intervalo de -0.5 a 0.5, 1 con el intervalo de 0.5 a $1.5, \ldots$ y 4 con el intervalo de 3.5 a 4.5, estamos, por así decirlo, "extendiendo" los valores de la variable aleatoria discreta dada sobre una escala continua.

$f(x)$

$\frac{6}{16}$

$\frac{4}{16}$

$\frac{1}{16}$

0 1 2 3 x

Número de caras

Figura 3.2 Histograma de probabilidades.

$f(x)$

$\frac{6}{16}$

$\frac{4}{16}$

$\frac{1}{16}$

0 1 2 3 4 x

Número de caras

Figura 3.3 Gráfica de barras.

Puesto que cada rectángulo del histograma de la figura 3.2 tiene ancho unitario, podríamos haber dicho que las *áreas* de los rectángulos, en vez de sus alturas, son iguales a las probabilidades correspondientes. Hay ciertas ventajas al identificar las áreas de los rectángulos con las probabilidades, por ejemplo, cuando deseamos aproximar la gráfica de una distribución de probabilidad discreta con una curva continua. Esto se puede hacer aun cuando los rectángulos de un histograma no tengan todos un ancho unitario mediante el ajuste de las alturas de los rectángulos o al modificar la escala vertical.

La gráfica de la figura 3.3 se llama **gráfica de barras**, pero también se conoce como histograma. Como en la figura 3.2, la altura de cada rectángulo, o barra, es igual a la probabilidad del valor correspondiente de la variable aleatoria, pero no hay pretensión de tener una escala horizontal continua. Algunas veces, como se muestra en la figura 3.4, usamos líneas (rectángulos sin ancho) en vez de los rectángulos, pero todavía nos referimos a las gráficas como histogramas.

Aunque hay diversas ocasiones en las que usaremos las gráficas en este texto, los histogramas y los diagramas de barra se usan principalmente en la estadística descriptiva para comunicar visualmente la información proporcionada por una distribución de probabilidad o una distribución de datos reales.

Hay muchos problemas en los que interesa conocer la probabilidad de que el valor de una variable aleatoria sea menor que o igual a algún número real x. Así, escribamos que la probabilidad de que X asuma un valor menor que o igual a x como $F(x) = P(X \leqq x)$ y refirámonos a esta función definida para todos los números reales x como la **función de distribución**, o la **distribución acumulativa**, de X.

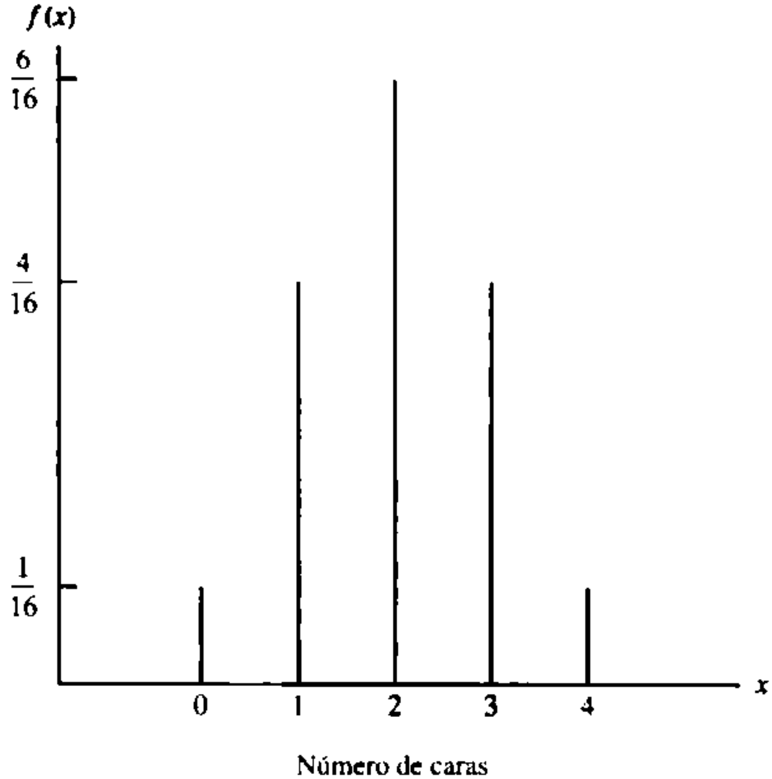

Figura 3.4 Histograma.

DEFINICIÓN 3.3 Si X es una variable aleatoria discreta, la función dada por

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t) \qquad \text{para } -\infty < x < \infty$$

donde $f(t)$ es el valor de la distribución de probabilidad de X en t, es llamada la **función de distribución**, o la **distribución acumulativa**, de X.

Con base en los postulados de probabilidad y algunas de sus consecuencias inmediatas, se sigue que

TEOREMA 3.2 Los valores de $F(x)$ de la función de distribución de una variable aleatoria discreta X satisface las condiciones

1. $F(-\infty) = 0$ y $F(\infty) = 1$;
2. si $a < b$, entonces $F(a) \leq F(b)$ para cualesquier números reales a y b.

Si se nos da la distribución de probabilidad de una variable aleatoria discreta, generalmente es fácil encontrar la función de distribución correspondiente.

EJEMPLO 3.5

Encuentre la función de distribución del número total de caras obtenidas en cuatro lanzamientos de una moneda balanceada.

Solución

Dado $f(0) = \frac{1}{16}$, $f(1) = \frac{4}{16}$, $f(2) = \frac{6}{16}$, $f(3) = \frac{4}{16}$ y $f(4) = \frac{1}{16}$ del ejemplo 3.3, se sigue que

$$F(0) = f(0) = \frac{1}{16}$$

$$F(1) = f(0) + f(1) = \frac{5}{16}$$

$$F(2) = f(0) + f(1) + f(2) = \frac{11}{16}$$

$$F(3) = f(0) + f(1) + f(2) + f(3) = \frac{15}{16}$$

$$F(4) = f(0) + f(1) + f(2) + f(3) + f(4) = 1$$

Por tanto, la función de distribución está dada por

$$F(x) = \begin{cases} 0 & \text{para } x < 0 \\ \frac{1}{16} & \text{para } 0 \leqq x < 1 \\ \frac{5}{16} & \text{para } 1 \leqq x < 2 \\ \frac{11}{16} & \text{para } 2 \leqq x < 3 \\ \frac{15}{16} & \text{para } 3 \leqq x < 4 \\ 1 & \text{para } x \geqq 4 \end{cases}$$

Observe que esta función de distribución está definida no sólo para los valores asumidos por la variable aleatoria dada, sino para todos los números reales. Por ejemplo, podemos escribir $F(1.7) = \frac{5}{16}$ y $F(100) = 1$, aunque las probabilidades de obtener "cuando mucho 1.7 caras" o "cuando mucho 100 caras" en cuatro lanzamientos de una moneda balanceada pueden no tener un significado real. ▲

EJEMPLO 3.6

Encuentre la función de distribución de la variable aleatoria W del ejemplo 3.1 y dibuje su gráfica.

Solución

Con base en las probabilidades dadas en la tabla de la página 75, podemos escribir $f(0) = \frac{3}{28}$, $f(1) = \frac{15}{56} + \frac{15}{56} = \frac{15}{28}$ y $f(2) = \frac{5}{14}$, de manera que

$$F(0) = f(0) = \frac{3}{28}$$

$$F(1) = f(0) + f(1) = \frac{9}{14}$$

$$F(2) = f(0) + f(1) + f(2) = 1$$

De ahí que la función de distribución de W está dada por

$$F(w) = \begin{cases} 0 & \text{para } w < 0 \\ \frac{3}{28} & \text{para } 0 \leqq w < 1 \\ \frac{9}{14} & \text{para } 1 \leqq w < 2 \\ 1 & \text{para } w \geqq 2 \end{cases}$$

La gráfica de esta función de distribución, mostrada en la figura 3.5, se obtuvo al trazar primero los puntos $(w, F(w))$ para $w = 0$, 1 y 2 y después completar la función escalón como se indica. Observe que en todos los puntos de discontinuidad la función de distribución asume el mayor de los dos valores, como lo indican los puntos grandes en la figura 3.5. ▲

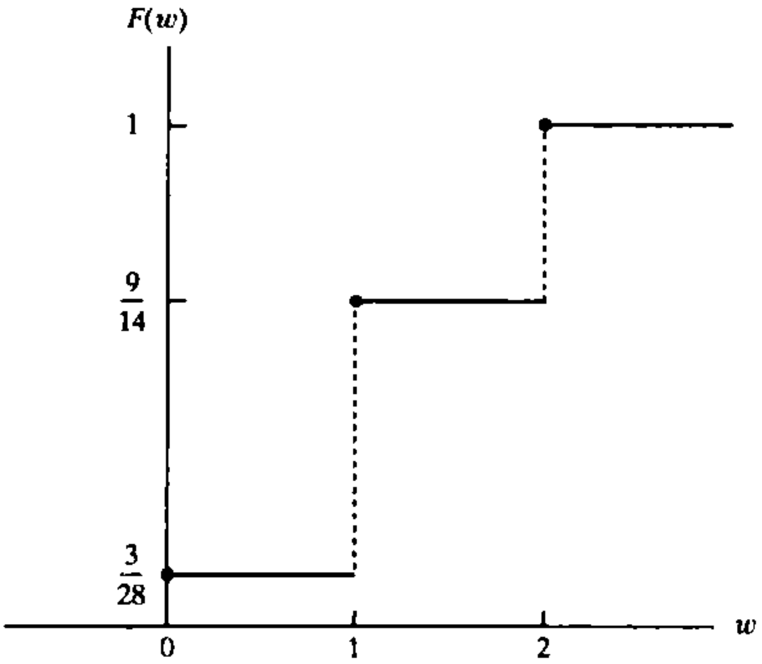

Figura 3.5 Gráfica de la función de distribución del ejemplo 3.6.

También podemos invertir el proceso ilustrado en los dos ejemplos anteriores, esto es, obtener valores de la distribución de probabilidad de una variable aleatoria a partir de su función de distribución. Para este fin, usamos el siguiente resultado:

TEOREMA 3.3 Si el intervalo de la variable aleatoria X consta de los valores $x_1 < x_2 < x_3 < \cdots < x_n$, entonces $f(x_1) = F(x_1)$ y

$$f(x_i) = F(x_i) - F(x_{i-1}) \qquad \text{para } i = 2, 3, \ldots, n$$

EJEMPLO 3.7

Si la función de distribución de X está dada por

$$F(x) = \begin{cases} 0 & \text{para } x < 2 \\ \dfrac{1}{36} & \text{para } 2 \leqq x < 3 \\ \dfrac{3}{36} & \text{para } 3 \leqq x < 4 \\ \dfrac{6}{36} & \text{para } 4 \leqq x < 5 \\ \dfrac{10}{36} & \text{para } 5 \leqq x < 6 \\ \dfrac{15}{36} & \text{para } 6 \leqq x < 7 \\ \dfrac{21}{36} & \text{para } 7 \leqq x < 8 \\ \dfrac{26}{36} & \text{para } 8 \leqq x < 9 \\ \dfrac{30}{36} & \text{para } 9 \leqq x < 10 \\ \dfrac{33}{36} & \text{para } 10 \leqq x < 11 \\ \dfrac{35}{36} & \text{para } 11 \leqq x < 12 \\ 1 & \text{para } x \geqq 12 \end{cases}$$

encuentre la distribución de probabilidad de esta variable aleatoria.

Solución

Al usar el teorema 3.3, obtenemos $f(2) = \frac{1}{36}$, $f(3) = \frac{3}{36} - \frac{1}{36} = \frac{2}{36}$, $f(4) = \frac{6}{36} - \frac{3}{36} = \frac{3}{36}$, $f(5) = \frac{10}{36} - \frac{6}{36} = \frac{4}{36}, \ldots,$ $f(12) = 1 - \frac{35}{36} = \frac{1}{36}$ y una comparación con las probabilidades en la tabla de la página 77 nos revela que la variable aleatoria que nos interesa en este caso es el número total de puntos tirados con un par de dados. ▲

En el resto de este capítulo estaremos interesados en variables aleatorias continuas y sus distribuciones, y en problemas relacionados con la ocurrencia simultánea de los valores de dos o más variables aleatorias. En el capítulo 5 regresaremos a las distribuciones de probabilidad de variables aleatorias discretas; de hecho, todo ese capítulo estará dedicado a las distribuciones de probabilidad que proporcionan modelos especialmente importantes para las aplicaciones.

EJERCICIOS

3.1 Determine si los valores dados pueden servir como valores de una distribución de probabilidad de una variable aleatoria con el intervalo $x = 1, 2, 3$ y 4:

(a) $f(1) = 0.25, f(2) = 0.75, f(3) = 0.25$ y $f(4) = -0.25$;

(b) $f(1) = 0.15, f(2) = 0.27, f(3) = 0.29$ y $f(4) = 0.29$;

(c) $f(1) = \frac{1}{19}, f(2) = \frac{10}{19}, f(3) = \frac{2}{19}$ y $f(4) = \frac{5}{19}$.

3.2 Determine si la función dada puede servir como distribución de probabilidad de una variable aleatoria con el intervalo dado:

(a) $f(x) = \dfrac{x-2}{5}$ para $x = 1, 2, 3, 4, 5$;

(b) $f(x) = \dfrac{x^2}{30}$ para $x = 0, 1, 2, 3, 4$;

(c) $f(x) = \dfrac{1}{5}$ para $x = 0, 1, 2, 3, 4, 5$.

3.3 Verifique que $f(x) = \dfrac{2x}{k(k+1)}$ para $x = 1, 2, 3, \ldots, k$ pueda servir como la distribución de probabilidad de una variable aleatoria con el intervalo dado.

3.4 Determine c tal que la función pueda servir como la distribución de probabilidad de una variable aleatoria con el intervalo dado:

(a) $f(x) = cx$ para $x = 1, 2, 3, 4, 5$;

(b) $f(x) = c\dbinom{5}{x}$ para $x = 0, 1, 2, 3, 4, 5$;

(c) $f(x) = cx^2$ para $x = 1, 2, 3, \ldots, k$;

(d) $f(x) = c\left(\dfrac{1}{4}\right)^x$ para $x = 1, 2, 3, \ldots$.

[*Sugerencia*: para el inciso (c) refiérase al apéndice A.]

3.5 ¿Para qué valores de k

$$f(x) = (1-k)k^x$$

puede servir como los valores de la distribución de probabilidad de una variable aleatoria con el intervalo infinito numerable $x = 0, 1, 2, \ldots$?

3.6 Demuestre que no hay valores de c tales que

$$f(x) = \frac{c}{x}$$

pueda servir como los valores de la distribución de probabilidad de una variable aleatoria con el intervalo infinito numerable $x = 1, 2, 3, \ldots$.

3.7 Construya un histograma de probabilidad para cada una de las siguientes distribuciones de probabilidad:

(a) $f(x) = \dfrac{\binom{2}{x}\binom{4}{3-x}}{\binom{6}{3}}$ para $x = 0, 1, 2$;

(b) $f(x) = \binom{5}{x}\left(\frac{1}{5}\right)^x\left(\frac{4}{5}\right)^{5-x}$ para $x = 0, 1, 2, 3, 4, 5$.

3.8 Demuestre el teorema 3.2.

3.9 Determine si los valores dados pueden servir como los valores de una función de distribución de una variable aleatoria con el intervalo $x = 1, 2, 3$ y 4:

(a) $F(1) = 0.3$, $F(2) = 0.5$, $F(3) = 0.8$ y $F(4) = 1.2$;

(b) $F(1) = 0.5$, $F(2) = 0.4$, $F(3) = 0.7$ y $F(4) = 1.0$;

(c) $F(1) = 0.25$, $F(2) = 0.61$, $F(3) = 0.83$, y $F(4) = 1.0$.

3.10 Encuentre la función de distribución de la variable aleatoria del inciso (a) del ejercicio 3.7 y dibuje su gráfica.

3.11 Encuentre la función de distribución de la variable aleatoria que tiene la distribución de probabilidad

$$f(x) = \frac{x}{15} \quad \text{para } x = 1, 2, 3, 4, 5$$

3.12 Si X tiene la función de distribución

$$F(x) = \begin{cases} 0 & \text{para } x < 1 \\ \frac{1}{3} & \text{para } 1 \leqq x < 4 \\ \frac{1}{2} & \text{para } 4 \leqq x < 6 \\ \frac{5}{6} & \text{para } 6 \leqq x < 10 \\ 1 & \text{para } x \geqq 10 \end{cases}$$

encuentre

(a) $P(2 < X \leqq 6)$;

(b) $P(X = 4)$;

(c) la distribución de probabilidad de X.

3.13 Si X tiene la función de distribución

$$F(x) = \begin{cases} 0 & \text{para } x < -4 \\ \frac{1}{4} & \text{para } -1 \leqq x < 1 \\ \frac{1}{2} & \text{para } 1 \leqq x < 3 \\ \frac{3}{4} & \text{para } 3 \leqq x < 5 \\ 1 & \text{para } x \geqq 5 \end{cases}$$

encuentre

(a) $P(X \leqq 3)$; (b) $P(X = 3)$; (c) $P(X < 3)$;
(d) $P(X \geqq 1)$; (e) $P(-0.4 < X < 4)$; (f) $P(X = 5)$.

3.14 Con respecto al ejemplo 3.4, verifique que los valores de la función de distribución están dados por

$$F(x) = \frac{x^2 + 5x}{50}$$

para $x = 1, 2, 3, 4$ y 5.

3.15 Con respecto al teorema 3.3, verifique que

(a) $P(X > x_i) = 1 - F(x_i)$ para $i = 1, 2, 3, \ldots, n$;
(b) $P(X \geqq x_i) = 1 - F(x_{i-1})$ para $i = 2, 3, \ldots, n$ y $P(X \geqq x_1) = 1$.

APLICACIONES

3.16 Con respecto al ejemplo 3.3, encuentre la distribución de probabilidad de Y, la diferencia entre el número de caras y el número de cruces obtenidas en cuatro lanzamientos de una moneda balanceada.

3.17 Una urna contiene cuatro pelotas numeradas 1, 2, 3 y 4. Si se sacan al azar dos pelotas de la urna (esto es, cada par tiene la misma oportunidad de ser seleccionado) y Z es la suma de los números de las pelotas que se sacaron, encuentre

(a) la distribución de probabilidad de Z y dibuje un histograma;
(b) la función de distribución de Z y dibuje su gráfica.

3.18 Una moneda está alterada para que las caras sean doblemente probables que las cruces. Para tres lanzamientos independientes de la moneda, encuentre

(a) la distribución de probabilidad de X, el número total de caras;
(b) la probabilidad de lanzar cuando mucho dos caras.

3.19 Con respecto al ejercicio 3.18, encuentre la función de distribución de la variable aleatoria X y dibuje su gráfica. Use la función de distribución de X para encontrar

(a) $P(1 < X \leqq 3)$; (b) $P(X > 2)$.

3.20 La distribución de probabilidad de V, el número semanal de accidentes en una cierta intersección, está dado por $g(0) = 0.40$, $g(1) = 0.30$, $g(2) = 0.20$ y $g(3) = 0.10$. Construya la función de distribución de V y dibuje su gráfica.

3.21 Con referencia al ejercicio 3.20, encuentre la probabilidad de que habrá al menos dos accidentes en una semana cualquiera, usando
(a) las probabilidades originales;
(b) los valores de la función de distribución.

3.22 Con respecto al ejercicio 2.58, ¿es el resultado de este ejercicio una distribución de probabilidad? Si lo es, dibuje su histograma.

3.23 Con respecto al ejercicio 3.22, encuentre la función de distribución de la suma de los puntos del dado, esto es la probabilidad de que esta suma de puntos en los dados sea cuando mucho S, donde $S = 2, 3, \ldots, 12$.

3.3 VARIABLES ALEATORIAS CONTINUAS

En la sección 3.1 introdujimos el concepto de variable aleatoria como una función de valor real definida sobre los puntos de un espacio muestral con una medida de probabilidad, y en la figura 3.1 ilustramos esto al asignar el total obtenido con un par de dados a cada uno de los 36 puntos igualmente probables del espacio muestral. En el caso continuo, donde las variables aleatorias pueden asumir valores en una escala continua, el procedimiento es con mucho el mismo. Los resultados de los experimentos se representan por puntos en segmentos de líneas o líneas, y los valores de las variables aleatorias son números apropiadamente asignados a los puntos mediante reglas o ecuaciones. Cuando una medición u observación proporciona directamente el valor de una variable aleatoria, generalmente no nos molestamos en distinguir entre el valor de la variable aleatoria (la medición que obtenemos) y el resultado del experimento (el punto correspondiente sobre el eje real). Así, si un experimento consiste en determinar el contenido real de un frasco de 230 gramos de café instantáneo, el resultado en sí, digamos, de 225.3 gramos, es el valor de la variable aleatoria que nos interesa, y no es realmente necesario añadir que el espacio muestral consiste en un cierto intervalo continuo de puntos sobre el eje real positivo.

El problema de definir probabilidades en relación con los espacios muestra continuos y las variables aleatorias continuas acarrea algunas complicaciones. Para ilustrar, consideremos la siguiente situación.

EJEMPLO 3.8

Supongamos que nos interesa la posibilidad de que un accidente ocurra en una carretera que tiene 200 km de largo y que nos interesa la probabilidad de que ocurra en un lugar dado o en cierto tramo de la carretera. El espacio muestral de este "experimento" consiste en un continuo de puntos, aquéllos en el intervalo de 0 a 200, y supondremos, para esclarecer el tema, que la probabilidad de que un accidente ocurra en cualquier intervalo de longitud d es $\frac{d}{200}$, donde d se mide en kilómetros. Advierta que esta asignación de probabilidades es congruente con los postulados 1 y 2 de la página 37: las probabilidades $\frac{d}{200}$ son todas no negativas y $P(S) = \frac{200}{200} = 1$. Hasta ahora la asignación de probabilidades se aplica sólo a los intervalos sobre el segmento de línea desde 0 hasta 200, pero si usamos el postulado 3, también podemos obtener las probabilidades para la unión de cualquier secuencia finita o infinita numerable de intervalos que no se traslapan. Por ejemplo, la probabilidad de que un accidente ocurra en cualesquiera de dos intervalos que no se traslapan de longitud d_1 y d_2 es

$$\frac{d_1 + d_2}{200}$$

y la probabilidad de que ocurra en uno cualquiera de una secuencia infinita numerable de intervalos que no se traslapan de longitud $d_1, d_2, d_3, \ldots$ es

$$\frac{d_1 + d_2 + d_3 + \cdots}{200}$$

Entonces, si aplicamos el teorema 2.7, podemos ampliar la asignación de probabilidad a la unión de intervalos que se traslapan, y puesto que la intersección de dos intervalos es un intervalo y el complemento de un intervalo es o un intervalo o la unión de dos intervalos, podemos ampliar la asignación de probabilidad a cualquier subconjunto del espacio muestral que se pueda obtener al formar uniones o intersecciones de finitamente o numerablemente muchos intervalos o al formar complementos. ▲

Así, al ampliar el concepto de probabilidad al caso continuo, hemos usado otra vez los postulados 1, 2 y 3, pero para hacer esto en general debemos excluir de nuestra definición de "evento" todos los subconjuntos del espacio muestral que no se puedan obtener formando uniones o intersecciones de finitamente o numerablemente muchos intervalos o formando complementos. Hablando en términos prácticos, esto no es importante, ya que simplemente no asignamos probabilidades a tales clases de conjuntos de difícil comprensión.

Con respecto al ejemplo 3.8, observe también que la probabilidad de que ocurra un accidente en un intervalo muy corto, digamos, un intervalo de 1 centímetro, es solamente 0.00000005, que es muy pequeño. Conforme la longitud del intervalo se aproxima a cero, la probabilidad de que ocurra un accidente en él también se aproxima a cero; ciertamente, en el caso continuo siempre asignamos una probabilidad cero a puntos individuales. Esto no significa que los eventos correspondientes no pueden ocurrir; después de todo, cuando ocurre un accidente en el tramo de 200 kilómetros de carretera, tiene que ocurrir en algún punto aun cuando cada punto tenga probabilidad cero.

3.4 FUNCIONES DE DENSIDAD DE PROBABILIDADES

La manera en que asignamos probabilidades en el ejemplo 3.8 es muy especial, y es de naturaleza similar a la manera en que asignamos probabilidades iguales a las seis caras de un dado, caras o cruces, las 52 cartas de juego en una baraja estándar y así sucesivamente. Para tratar de manera más general el problema de asociar probabilidades con los valores de variables aleatorias continuas, supongamos que un embotellador de bebidas gaseosas está interesado en la cantidad real de una bebida gaseosa que su máquina de embotellar pone en las botellas de 16 onzas. Evidentemente, la cantidad variará algo de botella a botella; es, de hecho, una variable aleatoria continua. Sin embargo, si él redondea las cantidades al décimo de onza más cercano, estará tratando con una variable aleatoria discreta que tiene una distribución de probabilidad, y esta distribución de probabilidad se puede ver como un histograma en el cual las probabilidades están dadas por las áreas de los rectángulos, digamos, como el diagrama en la parte superior de la figura 3.6. Si él redondea las cantidades al centésimo de onza más cercano, estará otra vez tratando con una variable aleatoria discreta (una diferente) que tiene una distribución de probabilidad, y esta distribución de probabilidad se puede ver como un histograma donde las probabilidades están dadas por las áreas de los rectángulos, digamos, como el diagrama en la parte media de la figura 3.6.

15.9 16.0 16.1

Cantidades redondeadas al décimo de onza más cercano

16.00 16.10

Cantidades redondeadas al centésimo de onza más cercano

15.9 16.0 16.1

Figura 3.6 Definición de probabilidad en el caso continuo.

Debe ser evidente que si redondeamos las cantidades al milésimo de onza más cercano o al diezmilésimo de onza más cercano, los histogramas de las distribuciones de probabilidad de las variables aleatorias discretas correspondientes se aproximarán a la curva continua mostrada en el diagrama en la parte inferior de la figura 3.6, y la suma de las áreas de los rectángulos que representan la probabilidad de que la cantidad quede dentro de un intervalo especificado se aproxima al área correspondiente bajo la curva.

Ciertamente, la definición de probabilidad en el caso continuo presume para cada variable aleatoria la existencia de una función, llamada una **función de densidad de probabilidad**, tal que las áreas bajo la curva dan las probabilidades asociadas con los intervalos correspondientes sobre el eje horizontal. En otras palabras, una función de densidad de probabilidad, integrada de a a b (con $a \leqq b$), da la probabilidad de que la variable aleatoria correspondiente asumirá un valor en el intervalo de a a b.

DEFINICIÓN **3.4** Una función con valores $f(x)$, definida sobre el conjunto de todos los números reales, se llama una **función de densidad de probabilidad** de la variable aleatoria continua X si y sólo si

$$P(a \leqq X \leqq b) = \int_a^b f(x)\,dx$$

para cualesquiera constantes reales a y b con $a \leqq b$.

Las funciones de densidad de probabilidad también se conocen, más brevemente, como **densidades de probabilidad**, **funciones de densidad**, **densidades**, o **f.d.p.**

Advierta que $f(c)$, el valor de la densidad de probabilidad de X en c, no da $P(X = c)$ como en el caso discreto. En relación con la variables aleatorias continuas, las probabilidades siempre están asociadas con intervalos y $P(X = c) = 0$ para cualquier constante real c. Esto concuerda con lo que dijimos en la página 90 y también se sigue directamente de la definición 3.4 con $a = b = c$.

Debido a esta propiedad, el valor de una función de densidad de probabilidad se puede cambiar con algunos de los valores de una variable aleatoria sin cambiar las probabilidades, y esto es por lo que decimos en la definición 3.4 que $f(x)$ es el valor de *una* densidad de probabilidad, *no* la densidad de probabilidad, de la variable aleatoria X en x. También, en vista de esta propiedad, no importa si incluimos los puntos terminales del intervalo de a a b; simbólicamente,

TEOREMA **3.4** Si X es una variable aleatoria continua y a y b son constantes reales con $a \leqq b$, entonces

$$P(a \leqq X \leqq b) = P(a \leqq X < b) = P(a < X \leqq b) = P(a < X < b)$$

Análogo al teorema 3.1, enunciemos ahora las siguientes propiedades de las densidades de probabilidad que, una vez más, se siguen directamente de los postulados de probabilidad.

TEOREMA 3.5 Una función puede servir como una densidad de probabilidad de una variable aleatoria continua X si sus valores, $f(x)$, satisfacen las condiciones†

1. $f(x) \geqq 0$ para $-\infty < x < \infty$;
2. $\int_{-\infty}^{\infty} f(x)\,dx = 1$.

EJEMPLO 3.9

Si X tiene la densidad de probabilidad

$$f(x) = \begin{cases} k \cdot e^{-3x} & \text{para } x > 0 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

encuentre k y $P(0.5 \leqq X \leqq 1)$.

Solución

Para satisfacer la segunda condición del teorema 3.5, debemos tener

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{0}^{\infty} k \cdot e^{-3x}\,dx = k \cdot \lim_{t \to \infty} \left. \frac{e^{-3x}}{-3} \right|_{0}^{t} = \frac{k}{3} = 1$$

y se sigue que $k = 3$. Para la probabilidad obtenemos

$$P(0.5 \leqq X \leqq 1) = \int_{0.5}^{1} 3e^{-3x}\,dx = -e^{-3x}\Big|_{0.5}^{1} = -e^{-3} + e^{-1.5} = 0.173 \qquad \blacktriangle$$

Aunque la variable aleatoria del ejemplo anterior no puede asumir valores negativos, ampliamos artificialmente el dominio de su densidad de probabilidad para incluir todos los números reales. Ésta es una práctica que seguiremos en todo este texto.

Como en el caso discreto, hay muchos problemas en los que nos interesa conocer la probabilidad de que el valor de una variable aleatoria continua X sea menor que o igual a algún número real x. Así, hagamos la siguiente definición análoga a la definición 3.3.

† Las condiciones no son "si y sólo si" como en el teorema 3.1 porque $f(x)$ podría ser negativa para algunos valores de la variable aleatoria sin afectar a cualesquiera de las probabilidades. Sin embargo, ambas condiciones del teorema 3.5 las satisfacen casi todas las densidades de probabilidad usadas en la práctica y estudiadas en este texto.

DEFINICIÓN 3.5 Si X es una variable aleatoria continua y el valor de su densidad de probabilidad en t es $f(t)$, entonces la función dada por

$$F(x) = P(X \leqq x) = \int_{-\infty}^{x} f(t)\,dt \qquad \text{para } -\infty < x < \infty$$

se llama la **función de distribución**, o la **distribución acumulativa**, de X.

Las propiedades de las funciones de distribución dadas en el teorema 3.2 son también válidas en el caso continuo; esto es, $F(-\infty) = 0$, $F(\infty) = 1$ y $F(a) \leqq (b)$ cuando $a < b$. Además, se sigue directamente de la definición 3.5 que

TEOREMA 3.6 Si $f(x)$ y $F(x)$ son los valores de la densidad de probabilidad y la función de distribución de X en x, entonces

$$P(a \leqq X \leqq b) = F(b) - F(a)$$

para cualesquier constantes reales a y b con $a \leqq b$, y

$$f(x) = \frac{dF(x)}{dx}$$

donde la derivada existe.

EJEMPLO 3.10

Encuentre la función de distribución de la variable aleatoria X del ejemplo 3.9, y úsela para evaluar de nuevo $P(0.5 \leqq X \leqq 1)$.

Solución

Para $x > 0$,

$$F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_{0}^{x} 3e^{-3t}\,dt = -e^{-3t}\Big|_{0}^{x} = 1 - e^{-3x}$$

y puesto que $F(x) = 0$ para $x \leqq 0$, , podemos escribir

$$F(x) = \begin{cases} 0 & \text{para } x \leqq 0 \\ 1 - e^{-3x} & \text{para } x > 0 \end{cases}$$

Para determinar la probabilidad $P(0.5 \leqq X \leqq 1)$,usamos la primera parte del teorema 3.6, y obtenemos

$$\begin{aligned} P(0.5 \leqq X \leqq 1) &= F(1) - F(0.5) \\ &= (1 - e^{-3}) - (1 - e^{-1.5}) \\ &= 0.173 \end{aligned}$$

Esto concuerda con el resultado obtenido al usar la densidad de probabilidad directamente en el ejemplo 3.9. ▲

EJEMPLO 3.11

Encuentre una función de densidad de probabilidad para la variable aleatoria cuya función de distribución está dada por

$$F(x) = \begin{cases} 0 & \text{para } x \leqq 0 \\ x & \text{para } 0 < x < 1 \\ 1 & \text{para } x \geqq 1 \end{cases}$$

y dibuje su gráfica

Solución

Puesto que la función de densidad dada es diferenciable en todas partes excepto en $x = 0$ y $x = 1$, diferenciamos para $x < 0$, $0 < x < 1$ y $x > 1$, obteniendo 0, 1 y 0. Así, de acuerdo a la segunda parte del teorema 3.6, podemos escribir

$$f(x) = \begin{cases} 0 & \text{para } x < 0 \\ 1 & \text{para } 0 < x < 1 \\ 0 & \text{para } x > 1 \end{cases}$$

Para llenar los vacíos en $x = 0$ y $x = 1$, hacemos $f(0)$ y $f(1)$ ambos igual a cero. En realidad, no importa cómo se defina la densidad de probabilidad en estos dos puntos, pero hay ciertas ventajas (las cuales se explican en la página 243) para escoger las variables de manera que la densidad de probabilidad sea diferente de cero en un intervalo abierto. Así, podemos escribir la densidad de probabilidad de la variable aleatoria original como

$$f(x) = \begin{cases} 1 & \text{para } 0 < x < 1 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

Esta gráfica se muestra en la figura 3.7 ▲

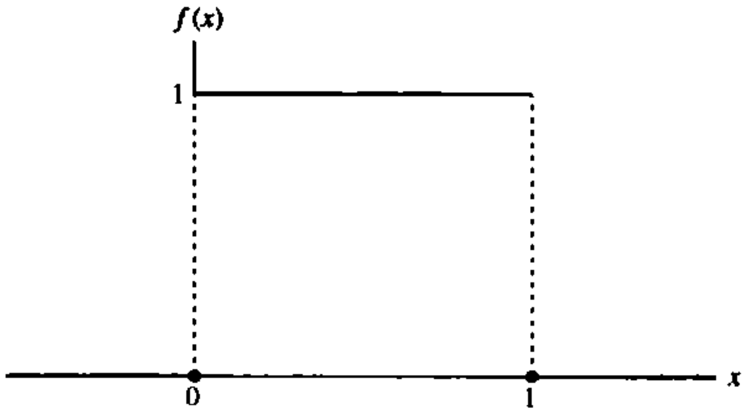

Figura 3.7 Densidad de probabilidad del ejemplo 3.11.

En la mayoría de las aplicaciones encontramos variables aleatorias que son discretas o son continuas, de manera que las funciones de distribución correspondientes tienen una apariencia de escalera como en la figura 3.5, son curvas continuas o son combinaciones de líneas como en la figura 3.8, que muestra la gráfica de la función de distribución del ejemplo 3.11.

Las funciones de distribución discontinuas como la mostrada en la figura 3.9 se presentan cuando las variables aleatorias son **mixtas**. Una función de distribución tal será discontinua en cada punto que tenga una probabilidad no cero y continua en cualquier otra parte. Como en el caso discreto, la altura del escalón en el punto de discontinuidad da la probabilidad que la variable aleatoria asumirá ese valor particular. Con respecto a la figura 3.9, $P(X = 0.5) = \frac{3}{4} - \frac{1}{4} = \frac{1}{2}$, pero fuera de eso la variable aleatoria es como una variable aleatoria continua.

En este libro nos limitaremos a variables aleatorias que son discretas o continuas en donde las últimas tienen funciones de distribución que son diferenciables en todos los valores de las variables aleatorias excepto en un conjunto finito de valores

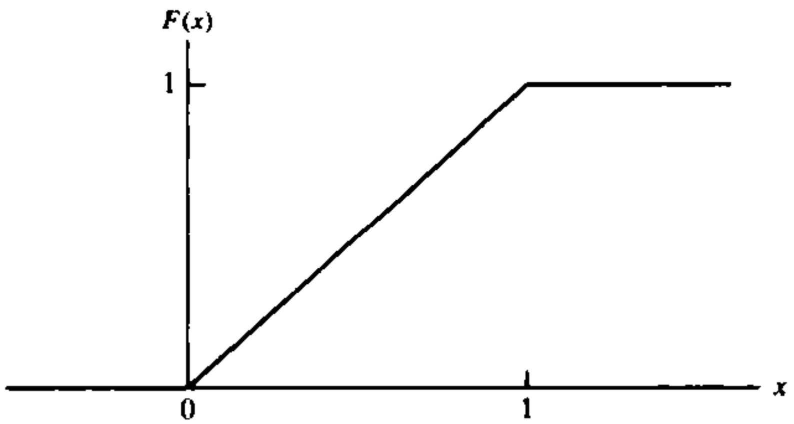

Figura 3.8 Función de distribución del ejemplo 3.11.

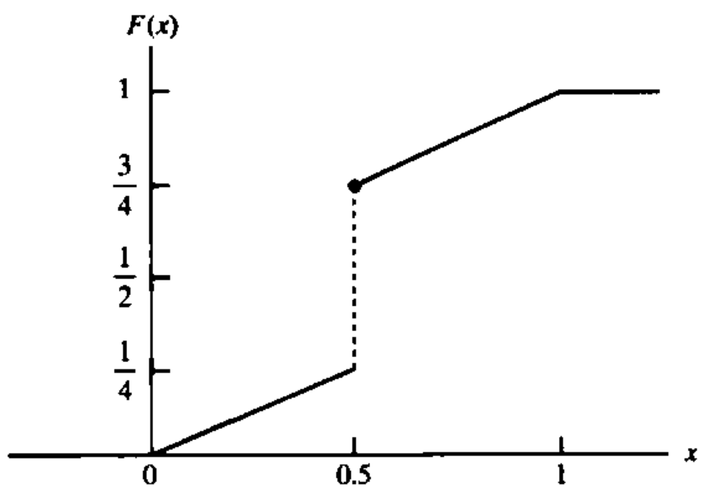

Figura 3.9 Función de distribución de una variable aleatoria mixta.

EJERCICIOS

3.24 La densidad de probabilidad de la variable aleatoria continua X está dada por

$$f(x) = \begin{cases} \dfrac{1}{5} & \text{para } 2 < x < 7 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

(a) dibuje su gráfica y verifique que el área total bajo la curva (arriba del eje x) es igual a 1.

(b) Encuentre $P(3 < X < 5)$.

3.25 Encuentre la función de distribución de la variable aleatoria X del ejercicio 3.24 y úsela para evaluar de nuevo el inciso (b).

3.26 (a) Demuestre que

$$f(x) = 3x^2 \qquad \text{para } 0 < x < 1$$

representa una función de densidad.

(b) Bosqueje una gráfica de esta función e indique el área asociada con la probabilidad que $0.1 < x < 0.5$.

(c) Calcule la probabilidad que $0.1 < x < 0.5$.

3.27 (a) Demuestre que

$$f(x) = e^{-x} \qquad \text{para } 0 < x < \infty$$

representa una función de densidad de probabilidad.

(b) Bosqueje una gráfica de esta función e indique el área asociada con la probabilidad de que $x > 1$.

(c) Calcule la probabilidad de que $x > 1$.

3.28 La densidad de probabilidad de la variable aleatoria Y está dada por

$$f(y) = \begin{cases} \dfrac{1}{8}(y + 1) & \text{para } 2 < y < 4 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

Encuentre $P(Y < 3.2)$ y $P(2.9 < Y < 3.2)$.

3.29 Encuentre la función de distribución de la variable aleatoria Y del ejercicio 3.28 y úsela para determinar las dos probabilidades pedidas en ese ejercicio.

3.30 La función de densidad de probabilidad de la variable aleatoria X está dada por

$$f(x) = \begin{cases} \dfrac{c}{\sqrt{x}} & \text{para } 0 < x < 4 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

Encuentre

(a) el valor de c;

(b) $P(X < \frac{1}{4})$ y $P(X > 1)$.

3.31 Encuentre la función de distribución de la variable aleatoria X del ejercicio 3.30 y úsela para determinar las dos probabilidades pedidas para el inciso (b) de ese ejercicio.

3.32 La densidad de probabilidad de la variable aleatoria Z está dada por

$$f(z) = \begin{cases} kze^{-z^2} & \text{para } z > 0 \\ 0 & \text{para } z \leqq 0 \end{cases}$$

Encuentre k y dibuje la gráfica de esta densidad de probabilidad.

3.33 Con referencia al ejercicio 3.32, encuentre la función de distribución de Z y dibuje su gráfica.

3.34 La función de densidad de la variable aleatoria X está dada por

$$g(x) = \begin{cases} 6x(1 - x) & \text{para } 0 < x < 1 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

Encuentre $P(X < \frac{1}{4})$ y $P(X > \frac{1}{2})$.

3.35 Con respecto al ejercicio 3.34, encuentre la función de distribución de X y úsela para evaluar de nuevo las dos probabilidades pedidas en ese ejercicio.

3.36 Encuentre la función de distribución de la variable aleatoria X cuya densidad de probabilidad está dada por

$$f(x) = \begin{cases} \dfrac{1}{3} & \text{para } 0 < x < 1 \\ \dfrac{1}{3} & \text{para } 2 < x < 4 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

También bosqueje las gráficas de la densidad de probabilidad y las funciones de distribución.

3.37 Encuentre la función de distribución de la variable aleatoria X cuya densidad de probabilidad está dada por

$$f(x) = \begin{cases} x & \text{para } 0 < x < 1 \\ 2 - x & \text{para } 1 \leqq x < 2 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

También bosqueje las gráficas de la densidad de probabilidad y las funciones de distribución.

3.38 Con respecto al ejercicio 3.37 encuentre $P(0.8 < X < 1.2)$, usando

(a) la densidad de probabilidad;

(b) la función de distribución.

3.39 Encuentre la función de distribución de la variable aleatoria X cuya densidad de probabilidad está dada por

$$f(x) = \begin{cases} \dfrac{x}{2} & \text{para } 0 < x \leqq 1 \\ \dfrac{1}{2} & \text{para } 1 < x \leqq 2 \\ \dfrac{3-x}{2} & \text{para } 2 < x < 3 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

También bosqueje las gráficas de estas densidades de probabilidad y funciones de distribución.

3.40 La función de distribución de la variable aleatoria X está dada por

$$F(x) = \begin{cases} 0 & \text{para } x < -1 \\ \dfrac{x+1}{2} & \text{para } -1 \leqq x < 1 \\ 1 & \text{para } x \geqq 1 \end{cases}$$

Encuentre $P(-\frac{1}{2} < X < \frac{1}{2})$ y $P(2 < X < 3)$.

3.41 Con referencia al ejercicio 3.40, encuentre la densidad de probabilidad de X y úsela para volver a calcular las dos probabilidades.

3.42 La función de distribución de la variable aleatoria Y está dada por

$$F(y) = \begin{cases} 1 - \dfrac{9}{y^2} & \text{para } y > 3 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

Encuentre $P(Y \leqq 5)$ y $P(Y > 8)$.

3.43 Con respecto al ejercicio 3.42, encuentre la densidad de probabilidad de Y y úsela para volver a calcular las dos probabilidades.

3.44 Con respecto al ejercicio 3.42 y el resultado del ejercicio 3.43, bosqueje las gráficas de la función de distribución y la densidad de probabilidad de Y, haciendo $f(3) = 0$.

3.45 La función de distribución de la variable aleatoria X está dada por

$$F(x) = \begin{cases} 1 - (1 + x)e^{-x} & \text{para } x > 0 \\ 0 & \text{para } x \leqq 0 \end{cases}$$

Encuentre $P(X \leqq 2)$, $P(1 < X < 3)$, y $P(X > 4)$.

3.46 Con respecto al ejercicio 3.45, encuentre la densidad de probabilidad de X.

3.47 Con respecto a la figura 3.9, encuentre expresiones para los valores de la función de distribución de la variable aleatoria mixta X para

(a) $x \leqq 0$; (b) $0 < x < 0.5$;
(c) $0.5 \leqq x < 1$; (d) $x \geqq 1$.

3.48 Use los resultados del ejercicio 3.47 para encontrar expresiones para los valores de la densidad de probabilidad de la variable aleatoria mixta X para

(a) $x < 0$; (b) $0 < x < 0.5$;
(c) $0.5 < x < 1$; (d) $x > 1$.

$P(X = 0.5) = \frac{1}{2}$, como ya se indicó en la página 96, y $f(0)$ y $f(1)$ están indefinidas.

3.49 La función de distribución de la variable aleatoria mixta Z está dada por

$$F(z) = \begin{cases} 0 & \text{para } z < -2 \\ \dfrac{z+4}{8} & \text{para } -2 \leqq z < 2 \\ 1 & \text{para } z \geqq 2 \end{cases}$$

Encuentre $P(Z = -2)$, $P(Z = 2)$, $P(-2 < Z < 1)$, y $P(0 \leqq Z \leqq 2)$.

APLICACIONES

3.50 La cantidad real de café (en gramos) en un frasco de 230 gramos que llena cierta máquina es una variable aleatoria cuya densidad de probabilidad está dada por

$$f(x) = \begin{cases} 0 & \text{para } x \leqq 227.5 \\ \dfrac{1}{5} & \text{para } 227.5 < x < 232.5 \\ 0 & \text{para } x \geqq 232.5 \end{cases}$$

Encuentre las probabilidades de que un frasco de 230 gramos que llene esta máquina contendrá

(a) cuando mucho 228.65 gramos de café;
(b) cualquier cantidad entre 229.34 y 231.66 gramos de café;
(c) al menos 229.85 gramos de café.

3.51 El número de minutos que un vuelo de Phoenix a Tucson se adelanta o atrasa es una variable aleatoria cuya densidad de probabilidad está dada por

$$f(x) = \begin{cases} \dfrac{1}{288}(36 - x^2) & \text{para } -6 < x < 6 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

donde los valores negativos son indicativos de que un vuelo llega adelantado y los valores positivos son indicativos de que llega con retraso. Encuentre las probabilidades de uno de estos vuelos llegará

(a) al menos 2 minutos adelantado;
(b) al menos 1 minuto retrasado;
(c) cualquier tiempo entre 1 y 3 minutos adelantado;
(d) exactamente 5 minutos retrasado.

3.52 La vida en el anaquel (en horas) de un alimento empacado perecedero es una variable aleatoria cuya función de densidad de probabilidad está dada por

$$f(x) = \begin{cases} \dfrac{20{,}000}{(x+100)^3} & \text{para } x > 0 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

Encuentre las probabilidades que uno de estos paquetes tendrá una vida en el anaquel de

(a) al menos 200 horas;

(b) cuando mucho 100 horas;

(c) cualquier tiempo entre 80 y 120 horas.

3.53 El desgaste (en miles de kilómetros) que los dueños de automóviles tienen con cierta clase de neumáticos es una variable aleatoria cuya densidad de probabilidad está dada por

$$f(x) = \begin{cases} \dfrac{1}{30} e^{-\frac{x}{30}} & \text{para } x > 0 \\ 0 & \text{para } x \leqq 0 \end{cases}$$

Encuentre la probabilidad que una de estos neumáticos durará

(a) cuando mucho 18,000 kilómetros;

(b) cualquier valor entre 27,000 y 36,000 kilómetros;

(c) al menos 48,000 kilómetros.

3.54 En una cierta ciudad el consumo diario de agua (en millones de litros) es una variable aleatoria cuya densidad de probabilidad está dada por

$$f(x) = \begin{cases} \dfrac{1}{9} x e^{-\frac{x}{3}} & \text{para } x > 0 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

¿Cuáles son las probabilidades de que en un día dado

(a) el consumo de agua en esta ciudad no sea mayor de 6 millones de litros;

(b) el abastecimiento de agua sea inadecuado si la capacidad diaria de esta ciudad es 9 millones de litros?

3.55 La duración de vida total (en años) de perros de cinco años de edad de una cierta raza es una variable aleatoria cuya función de distribución está dada por

$$F(x) = \begin{cases} 0 & \text{para } x \leqq 5 \\ 1 - \dfrac{25}{x^2} & \text{para } x > 5 \end{cases}$$

Encuentre las probabilidades de que un perro como ésos de cinco años de edad vivirá

(a) más allá de los 10 años;

(b) menos de 8 años;

(c) entre 12 y 15 años.

3.5 DISTRIBUCIONES MULTIVARIADAS

Al principio de este capítulo definimos una variable aleatoria como una función de valor real definida sobre un espacio muestral con una medida de probabilidad, y es lógico que se pueden definir muchas variables aleatorias diferentes sobre un y el mismo espacio muestral. Con respecto al espacio muestral de la figura 3.1, por ejemplo, sólo consideramos la variable aleatoria cuyos valores fueron los totales lanzados con un par de dados, pero también podíamos haber considerado la variable aleatoria cuyos valores fueran los productos de los números lanzados con los dos dados, la variable aleatoria cuyos valores sean la diferencia de los números lanzados con el dado rojo y el dado verde, la variable aleatoria cuyos valores sean 0, 1 o 2 dependiendo del número de dados que salieran con 2, y así sucesivamente. Más real, un experimento puede consistir en escoger aleatoriamente algunos de los 345 estudiantes que asisten a una escuela primaria, y tal vez al director le interesen sus coeficientes de inteligencia, a la enfermera de la escuela sus pesos, a sus profesores el número de días que han estado ausentes, y así sucesivamente.

En esta sección examinaremos primero en el caso **bivariado**, esto es, en las situaciones donde nos interesa, al mismo tiempo, un par de variables aleatorias definidas en un espacio muestral conjunto. Más tarde, ampliaremos este examen al caso **multivariado** que abarca cualquier número finito de variables aleatorias.

Si X y Y son variables aleatorias discretas, escribimos como $P(X = x, Y = y)$, la probabilidad que X asumirá el valor x y Y asumirá el valor y. Así, $P(X = x, Y = y)$ es la probabilidad de la intersección de los eventos $X = x$ y $Y = y$. Como en el **caso univariado**, donde tratamos con una variable aleatoria y podíamos mostrar las probabilidades asociadas con todos los valores de X mediante una tabla, podemos ahora, en el caso bivariado, mostrar las probabilidades asociadas con todos los pares de valores de X y Y mediante una tabla.

EJEMPLO 3.12

Dos cápsulas se seleccionan al azar de un frasco que contiene tres aspirinas, dos sedantes y cuatro cápsulas laxantes. Si X y Y son, respectivamente, los números de cápsulas de aspirina y sedantes incluidas entre las dos cápsulas que se sacaron del frasco, encuentre las probabilidades asociadas con todos los pares posibles de valores de X y Y.

Solución

Los pares posibles son $(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$, $(1, 2)$ y $(2, 0)$. Para encontrar la probabilidad asociada con $(1, 0)$, por ejemplo, observe que estamos interesados en el evento de obtener una de las tres cápsulas de aspirina, ninguna de las dos cápsulas de sedante y, por tanto, una de las cuatro cápsulas laxantes. El número de maneras en que esto se puede hacer es $\binom{3}{1}\binom{2}{0}\binom{4}{1} = 12$, y el número total de maneras en las que se pueden seleccionar dos de las nueve cápsulas es $\binom{9}{2} = 36$. Puesto que esas posibilidades son igualmente probables en virtud de la suposición de que la selección es al azar, se sigue del teorema 2.2 que la probabilidad asociada con (1, 0) es (1, 0) es $\frac{12}{36} = \frac{1}{3}$. Similarmente, la probabilidad asociada con (1, 1) es

$$\frac{\binom{3}{1}\binom{2}{1}\binom{4}{0}}{36} = \frac{6}{36} = \frac{1}{6}$$

y, continuando de esta manera, obtenemos los valores mostrados en la siguiente tabla

		x = 0	x = 1	x = 2
	0	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{12}$
y	1	$\frac{2}{9}$	$\frac{1}{6}$	
	2	$\frac{1}{36}$		

▲

En realidad, como en el caso univariado, generalmente es preferible representar probabilidades como éstas mediante una fórmula. En otras palabras, es preferible expresar las probabilidades mediante una función con los valores $f(x, y) = P(X = x, Y = y)$ para cualquier par de valores (x, y) dentro del intervalo de las variables aleatorias X y Y. Por ejemplo, veremos en el capítulo 5 que para los dos variables aleatorias del ejemplo 3.12 podemos escribir

$$f(x, y) = \frac{\binom{3}{x}\binom{2}{y}\binom{4}{2-x-y}}{\binom{9}{2}} \quad \text{para } x = 0, 1, 2; \quad y = 0, 1, 2; \quad 0 \leqq x + y \leqq 2$$

DEFINICIÓN 3.6 Si X y Y son variables aleatorias discretas, la función dada para $f(x, y) = P(X = x, Y = y)$ para cada par de valores (x, y) dentro del intervalo de (x, y) se llama **distribución de probabilidad conjunta** de X y Y.

Análogo al teorema 3.1, se sigue de los postulados de probabilidad que

TEOREMA 3.7 Una función bivariada puede servir como la distribución de probabilidad conjunta de un par de variables aleatorias discretas X y Y si y sólo si sus valores, $f(x, y)$, satisfacen las condiciones

1. $f(x, y) \geqq 0$ para cada par de valores de (x, y) dentro de su dominio.
2. $\sum_x \sum_y f(x, y) = 1$, donde la doble suma se extiende sobre todos los posibles pares de (x, y) dentro de su dominio.

EJEMPLO 3.13

Determine el valor de k para el que la función dada por

$$f(x, y) = kxy \qquad \text{para } x = 1, 2, 3; \quad y = 1, 2, 3$$

puede servir como una distribución de probabilidad conjunta.

Solución

Al sustituir los diversos valores de x y y, obtenemos $f(1, 1) = k$, $f(1, 2) = 2k$, $f(1, 3) = 3k$, $f(2, 1) = 2k$, $f(2, 2) = 4k$, $f(2, 3) = 6k$, $f(3, 1) = 3k$, $f(3, 2) = 6k$ y $f(3, 3) = 9k$. Para satisfacer la primera condición del teorema 3.7, la constante k debe ser no negativa, y para satisfacer la segunda condición,

$$k + 2k + 3k + 2k + 4k + 6k + 3k + 6k + 9k = 1$$

de manera que $36k = 1$ y $k = \frac{1}{36}$. ▲

Como en el caso univariado, hay muchos problemas en los que interesa conocer la probabilidad de que los valores de dos variables aleatorias sean menores que, o iguales a, algunos números reales x y y.

DEFINICIÓN 3.7 Si X y Y son variables aleatorias discretas, la función dada por

$$F(x, y) = P(X \leqq x, Y \leqq y) = \sum_{s \leqq x} \sum_{t \leqq y} f(s, t) \qquad \text{para } -\infty < x < \infty, \; -\infty < y < \infty$$

donde $f(s, t)$ es el valor de la distribución de probabilidad conjunta de X y Y en (s, t), se llama la **función de distribución conjunta**, o la **distribución acumulativa conjunta**, de X y Y.

En el ejercicio 3.62 se pedirá al lector que demuestre las propiedades de las funciones de distribución conjunta que son análogas a las del teorema 3.2.

EJEMPLO 3.14

Con respecto al ejemplo 3.12, encuentre $F(1, 1)$.

Solución

$$\begin{aligned} F(1, 1) &= P(X \leqq 1, Y \leqq 1) \\ &= f(0, 0) + f(0, 1) + f(1, 0) + f(1, 1) \\ &= \frac{1}{6} + \frac{2}{9} + \frac{1}{3} + \frac{1}{6} \\ &= \frac{8}{9} \qquad \blacktriangle \end{aligned}$$

Como en el caso univariado, la función de distribución conjunta de dos variables aleatorias se define para todos los números reales. Por ejemplo, para el ejemplo 3.12 también obtenemos $F(-2, 1) = P(X \leqq -2, Y \leqq 1) = 0$ y $F(3.7, 4.5) = P(X \leqq 3.7, Y \leqq 4.5) = 1$.

Ampliemos ahora los diversos conceptos introducidos en esta sección al caso continuo.

DEFINICIÓN **3.8** Una función bivariada con valores $f(x, y)$, definida sobre el plano xy, se llama **función de densidad de probabilidad conjunta** de las variables aleatorias continuas X y Y si y sólo si

$$P[(X, Y) \in A] = \int_A \int f(x, y)\, dx\, dy$$

para cualquier región A en el plano xy.

Análogo al teorema 3.5, se sigue de los postulados de probabilidad que

TEOREMA **3.8** Una función bivariada puede servir como una función de densidad de probabilidad conjunta de un par de variables aleatorias continuas X y Y si sus valores $f(x, y)$, satisfacen las condiciones

1. $f(x, y) \geqq 0 \qquad$ para $-\infty < x < \infty, \quad -\infty < y < \infty$;
2. $\displaystyle\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dx\, dy = 1$.

EJEMPLO 3.15

Dada la función de densidad de probabilidad conjunta

$$f(x, y) = \begin{cases} \frac{3}{5}x(y + x) & \text{para } 0 < x < 1, 0 < y < 2 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

de dos variables aleatorias X y Y, encuentre $P[(X, Y) \in A]$, donde A es la región $\{(x, y) | 0 < x < \frac{1}{2}, 1 < y < 2\}$.

Solución

$$\begin{aligned} P[(X, Y) \in A] &= P\left(0 < X < \frac{1}{2}, 1 < Y < 2\right) \\ &= \int_1^2 \int_0^{\frac{1}{2}} \frac{3}{5} x(y + x)\, dx\, dy \\ &= \int_1^2 \frac{3x^2y}{10} + \frac{3x^3}{15}\bigg|_{x=0}^{x=\frac{1}{2}} dy \\ &= \int_1^2 \left(\frac{3y}{40} + \frac{1}{40}\right) dy = \frac{3y^2}{80} + \frac{y}{40}\bigg|_1^2 \\ &= \frac{11}{80} \qquad \blacktriangle \end{aligned}$$

Análogo a la definición 3.7, tenemos la siguiente definición de la función de distribución conjunta de dos variables aleatorias continuas.

> DEFINICIÓN 3.9 Si X y Y son variables aleatorias continuas, la función dada por
>
> $$F(x, y) = P(X \leq x, Y \leq y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(s, t)\, ds\, dt \quad \text{para } -\infty < x < \infty, \; -\infty < y < \infty$$
>
> donde $f(s, t)$ es el valor de la densidad de probabilidad conjunta de X y Y en (s, t), se llama la **función de distribución conjunta** de X y Y.

Advierta que las propiedades de las funciones de distribución conjunta, que se pedirá al lector que demuestre en el ejercicio 3.62 para el caso discreto, también son válidas para el caso continuo.

Como en la sección 3.4, limitaremos nuestro examen aquí a las variables aleatorias cuya función de distribución conjunta es continua en todas partes y parcialmente diferenciable con respecto a cada variable para todos los valores de las dos variables aleatorias, excepto un conjunto finito de valores.

Análogo a la relación $f(x) = \frac{dF(x)}{dx}$ del teorema 3.6, la diferenciación parcial en la definición 3.9 nos lleva a

$$f(x, y) = \frac{\partial^2}{\partial x\, \partial y} F(x, y)$$

en todas partes donde existan estas derivadas parciales. Como en la sección 3.4, la función de distribución conjunta de dos variables aleatorias continuas determina su **densidad conjunta** (abreviación de función de densidad de probabilidad conjunta) en todos los puntos (x, y) donde la densidad conjunta es continua. También como en la sección 3.4, generalmente hacemos los valores de las densidades de probabilidad conjunta igual a cero siempre que no estén definidas por la relación anterior.

EJEMPLO 3.16

Si la densidad de probabilidad conjunta de X y Y está dada por

$$f(x, y) = \begin{cases} x + y & \text{para } 0 < x < 1, 0 < y < 1 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

encuentre la función de distribución conjunta de estas dos variables aleatorias.

Solución

Si $x < 0$ o $y < 0$, se sigue inmediatamente que $F(x, y) = 0$. Para $0 < x < 1$ y $0 < y < 1$ (Región I de la figura 3.10), obtenemos

$$F(x, y) = \int_0^y \int_0^x (s + t)\, ds\, dt = \frac{1}{2} xy(x + y)$$

para $x > 1$ y $0 < y < 1$ (Región II de la figura 3.10), obtenemos

$$F(x, y) = \int_0^y \int_0^1 (s + t)\, ds\, dt = \frac{1}{2} y(y + 1)$$

para $0 < x < 1$ y $y > 1$ (Región III de la figura 3.10), obtenemos

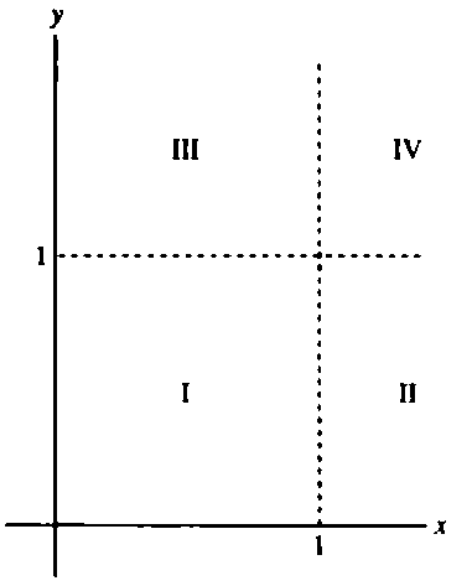

Figura 3.10 Diagrama para el ejemplo 3.16.

$$F(x, y) = \int_0^1 \int_0^x (s + t)\,ds\,dt = \frac{1}{2}x(x + 1)$$

y para $x > 1$ y $y > 1$ (Región IV de la figura 3.10), obtenemos

$$F(x, y) = \int_0^1 \int_0^1 (s + t)\,ds\,dt = 1$$

Puesto que la función de distribución conjunta es continua en todas partes, los límites entre dos regiones cualquiera se pueden incluir en cualquiera de los dos y podemos escribir

$$F(x, y) = \begin{cases} 0 & \text{para } x \leqq 0 \text{ o } y \leqq 0 \\ \frac{1}{2}xy(x + y) & \text{para } 0 < x < 1, 0 < y < 1 \\ \frac{1}{2}y(y + 1) & \text{para } x \geqq 1, 0 < y < 1 \\ \frac{1}{2}x(x + 1) & \text{para } 0 < x < 1, y \geqq 1 \\ 1 & \text{para } x \geqq 1, y \geqq 1 \end{cases}$$ ▲

EJEMPLO 3.17

Encuentre la densidad de probabilidad conjunta de las dos variables aleatorias X y Y cuya función de distribución conjunta está dada por

$$F(x, y) = \begin{cases} (1 - e^{-x})(1 - e^{-y}) & \text{para } x > 0 \text{ y } y > 0 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

También use la densidad de probabilidad conjunta para determinar

$$P(1 < X < 3, 1 < Y < 2).$$

Solución

Puesto que la diferenciación parcial nos da

$$\frac{\partial^2}{\partial x\,\partial y} F(x, y) = e^{-(x+y)}$$

para $x > 0$ y $y > 0$ y 0 en cualquier otra parte, encontramos que la densidad de probabilidad conjunta de X y Y está dada por

$$f(x, y) = \begin{cases} e^{-(x+y)} & \text{para } x > 0 \text{ y } y > 0 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

Así, la integración nos da

$$\int_1^2 \int_1^3 e^{-(x+y)}\, dx\, dy = (e^{-1} - e^{-3})(e^{-1} - e^{-2})$$
$$= e^{-2} - e^{-3} - e^{-4} + e^{-5}$$
$$= 0.074$$

para $P(1 < X < 3, 1 < Y < 1)$. ▲

Para dos variables aleatorias, la probabilidad conjunta es, geométricamente hablando, una superficie y la probabilidad que calculamos en el ejemplo precedente está dada por el volumen bajo esta superficie, como se muestra en la figura 3.11.

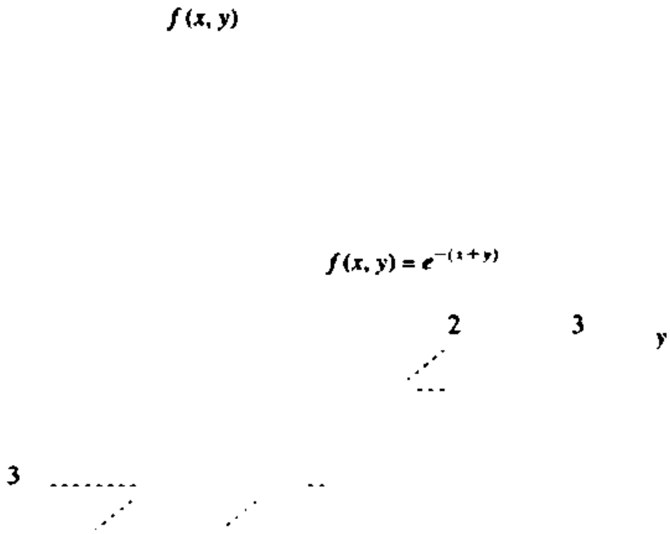

Figura 3.11 Diagrama para el ejemplo 3.17.

Todas las definiciones de esta sección se pueden generalizar al caso **multivariado**, donde hay n variables aleatorias. En correspondencia a la definición 3.6, los valores de distribución de probabilidad conjunta de n variables aleatorias discretas $X_1, X_2, \ldots$ y X_n está dado por

$$f(x_1, x_2, \ldots, x_n) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$$

para cada "n-ada" $(x_1, x_2, \ldots, x_n)$ dentro del intervalo de las variables aleatorias; y en correspondencia a la definición 3.7, los valores de su función de distribución conjunta están dados por

$$F(x_1, x_2, \ldots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n)$$

para $-\infty < x_1 < \infty, -\infty < x_2 < \infty, \ldots, -\infty < x_n < \infty$.

EJEMPLO 3.18

Si la distribución de probabilidad conjunta de tres variables aleatorias discretas X, Y y Z está dada por

$$f(x, y, z) = \frac{(x + y)z}{63} \qquad \text{para } x = 1, 2; \quad y = 1, 2, 3; \quad z = 1, 2$$

encuentre $P(X = 2, Y + Z \leqq 3)$.

Solución

$$\begin{aligned} P(X = 2, Y + Z \leqq 3) &= f(2, 1, 1) + f(2, 1, 2) + f(2, 2, 1) \\ &= \frac{3}{63} + \frac{6}{63} + \frac{4}{63} \\ &= \frac{13}{63} \qquad \blacktriangle \end{aligned}$$

En el caso continuo, se obtienen otra vez las probabilidades al integrar la densidad de probabilidad conjunta, y la función de distribución conjunta está dada por

$$F(x_1, x_2, \ldots, x_n) = \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} f(t_1, t_2, \ldots, t_n)\, dt_1\, dt_2 \cdots dt_n$$

para $-\infty < x_1 < \infty$, $-\infty < x_2 < \infty, \ldots, -\infty < x_n < \infty$, análoga a la definición 3.9. También, la diferenciación parcial nos da

$$f(x_1, x_2, \ldots, x_n) = \frac{\partial^n}{\partial x_1\, \partial x_2 \cdots \partial x_n} F(x_1, x_2, \ldots, x_n)$$

en cualquier parte en que estas derivadas parciales existan.

EJEMPLO 3.19

Si la densidad de probabilidad **trivariada** de X_1, X_2 y X_3 está dada por

$$f(x_1, x_2, x_3) = \begin{cases} (x_1 + x_2)e^{-x_3} & \text{para } 0 < x_1 < 1, 0 < x_2 < 1, x_3 > 0 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

encuentre $P[(X_1, X_2, X_3) \in A]$, donde A es la región

$$\left\{(x_1, x_2, x_3) | 0 < x_1 < \frac{1}{2}, \frac{1}{2} < x_2 < 1, x_3 < 1\right\}$$

Solución

$$P[(X_1, X_2, X_3) \in A] = P\left(0 < X_1 < \frac{1}{2}, \frac{1}{2} < X_2 < 1, X_3 < 1\right)$$

$$= \int_0^1 \int_{\frac{1}{2}}^1 \int_0^{\frac{1}{2}} (x_1 + x_2)e^{-x_3}\, dx_1\, dx_2\, dx_3$$

$$= \int_0^1 \int_{\frac{1}{2}}^1 \left(\frac{1}{8} + \frac{x_2}{2}\right)e^{-x_3}\, dx_2\, dx_3$$

$$= \int_0^1 \frac{1}{4} e^{-x_3}\, dx_3$$

$$= \frac{1}{4}(1 - e^{-1}) = 0.158 \qquad \blacktriangle$$

EJERCICIOS

3.56 Si los valores de la distribución de probabilidad conjunta de X y Y son como se muestra en la tabla

		x = 0	1	2
y	0	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{24}$
	1	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{40}$
	2	$\frac{1}{8}$	$\frac{1}{20}$	
	3	$\frac{1}{120}$		

encuentre

(a) $P(X = 1, Y = 2)$; (b) $P(X = 0, 1 \leqq Y < 3)$;
(c) $P(X + Y \leqq 1)$; (d) $P(X > Y)$.

3.57 Con respecto al ejercicio 3.56, encuentre los siguientes valores de la función de distribución conjunta de las dos variables aleatorias:

(a) $F(1.2, 0.9)$; (b) $F(-3, 1.5)$;
(c) $F(2, 0)$; (d) $F(4, 2.7)$.

3.58 Si la distribución de probabilidad conjunta de X y Y está dada por

$$f(x, y) = c(x^2 + y^2) \qquad \text{para } x = -1, 0, 1, 3; \quad y = -1, 2, 3$$

encuentre el valor de c.

3.59 Con respecto al ejercicio 3.58 y al valor obtenido para c, encuentre

(a) $P(X \leqq 1, Y > 2)$;

(b) $P(X = 0, Y \leqq 2)$;

(c) $P(X + Y > 2)$.

3.60 Demuestre que no hay un valor de k para el cual

$$f(x, y) = ky(2y - x) \qquad \text{para } x = 0, 3; \quad y = 0, 1, 2$$

pueda servir como la distribución de probabilidad conjunta de dos variables aleatorias.

3.61 Si la distribución de probabilidad conjunta de X y Y está dada por

$$f(x, y) = \frac{1}{30}(x + y) \qquad \text{para } x = 0, 1, 2, 3; \quad y = 0, 1, 2$$

elabore una tabla que muestre los valores de la función de distribución conjunta de las dos variables aleatorias en los 12 puntos $(0, 0), (0, 1), \ldots, (3, 2)$.

3.62 Si $F(x, y)$ es el valor de la función de distribución conjunta de dos variables aleatorias discretas X y Y en (x, y), demostrar que

(a) $F(-\infty, -\infty) = 0$;

(b) $F(\infty, \infty) = 1$;

(c) si $a < b$ y $c < d$, entonces $F(a, c) \leqq F(b, d)$.

3.63 Determine k de manera que

$$f(x, y) = \begin{cases} kx(x - y) & \text{para } 0 < x < 1, -x < y < x \\ 0 & \text{en cualquier otra parte} \end{cases}$$

pueda servir como una densidad de probabilidad conjunta.

3.64 Si la densidad de probabilidad conjunta de X y Y está dada por

$$f(x, y) = \begin{cases} 24xy & \text{para } 0 < x < 1, 0 < y < 1, x + y < 1 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

encuentre $P(X + Y < \frac{1}{2})$.

3.65 Si la densidad de probabilidad conjunta de X y Y está dada por

$$f(x, y) = \begin{cases} 2 & \text{para } x > 0, y > 0, x + y < 1 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

encuentre

(a) $P(X \leqq \frac{1}{2}, Y \leqq \frac{1}{2})$;

(b) $P(X + Y > \frac{2}{3})$;

(c) $P(X > 2Y)$.

3.66 Con referencia al ejercicio 3.65, encuentre una expresión para los valores de la función de distribución conjunta de X y Y cuando $x > 0$ $y > 0$ y $x + y < 1$, y úsela para verificar el resultado del inciso (a).

3.67 Si la densidad de probabilidad conjunta de X y Y está dada por

$$f(x, y) = \begin{cases} \frac{1}{y} & \text{para } 0 < x < y, 0 < y < 1 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

encuentre la probabilidad que la suma de los valores de X y Y excederá $\frac{1}{2}$.

3.68 Encuentre la densidad de probabilidad conjunta de las dos variables aleatorias X y Y cuya función de distribución conjunta está dada por

$$F(x, y) = \begin{cases} (1 - e^{-x^2})(1 - e^{-y^2}) & \text{para } x > 0, y > 0 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

3.69 Use la densidad de probabilidad conjunta obtenida en el ejercicio 3.68 para encontrar $P(1 < X \leqq 2, 1 < Y \leqq 2)$.

3.70 Encuentre la densidad de probabilidad conjunta de las dos variables aleatorias X y Y cuya función de distribución conjunta está dada por

$$F(x, y) = \begin{cases} 1 - e^{-x} - e^{-y} + e^{-x-y} & \text{para } x > 0, y > 0 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

3.71 Use la densidad de probabilidad conjunta obtenida en el ejercicio 3.70 para encontrar $P(X + Y > 3)$.

3.72 Si $F(x, y)$ es el valor de la función de distribución conjunta de las dos variables aleatorias continuas X y Y en (x, y), exprese $P(a < X \leqq b, c < Y \leqq d)$ en términos de $F(a, c)$, $F(a, d)$, $F(b, c)$ y $F(b, d)$. Observe que el resultado también es válido para variables aleatorias discretas.

3.73 Use la fórmula obtenida en el ejercicio 3.72 para verificar el resultado 0.074, del ejemplo 3.17.

3.74 Use la fórmula obtenida en el ejercicio 3.72 para verificar el resultado del ejercicio 3.69.

3.75 Use la fórmula obtenida en el ejercicio 3.72 para verificar el resultado del ejercicio 3.71.

3.76 Encuentre k si la densidad de probabilidad conjunta de X, Y y Z está dada por

$$f(x, y, z) = kxyz$$

para $x = 1, 2$; $y = 1, 2, 3$; $z = 1, 2$.

3.77 Con respecto al ejercicio 3.76, encuentre

(a) $P(X = 1, Y \leqq 2, Z = 1)$;

(a) $P(X = 2, Y + Z = 4)$.

3.78 Con referencia al ejercicio 3.76, encuentre los siguientes valores de la función de distribución conjunta de las tres variables aleatorias:

(a) $F(2, 1, 2)$;

(b) $F(1, 0, 1)$;

(c) $F(4, 4, 4)$.

3.79 Encuentre k si la densidad de probabilidad conjunta de X, Y y Z está dada por

$$f(x, y, z) = \begin{cases} kxy(1-z) & \text{para } 0 < x < 1, 0 < y < 1, 0 < z < 1, x + y + z < 1 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

3.80 Con respecto al ejercicio 3.79, encuentre $P(X + Y < \frac{1}{2})$.

3.81 Use el resultado del ejercicio 3.16 para verificar la función de distribución conjunta de las variables aleatorias X_1, X_2 y X_3 del ejemplo 3.19 está dada por

$$F(x_1, x_2, x_3) = \begin{cases} 0 & \text{para } x_1 \leqq 0, x_2 \leqq 0, \text{o } x_3 \leqq 0 \\ \frac{1}{2}x_1x_2(x_1 + x_2)(1 - e^{-x_3}) & \text{para } 0 < x_1 < 1, 0 < x_2 < 1, x_3 > 0 \\ \frac{1}{2}x_2(x_2 + 1)(1 - e^{-x_3}) & \text{para } x_1 \geqq 1, 0 < x_2 < 1, x_3 > 0 \\ \frac{1}{2}x_1(x_1 + 1)(1 - e^{-x_3}) & \text{para } 0 < x_1 < 1, x_2 \geqq 1, x_3 > 0 \\ 1 - e^{-x_3} & \text{para } x_1 \geqq 1, x_2 \geqq 1, x_3 > 0 \end{cases}$$

3.82 Si la densidad de probabilidad conjunta de X, Y y Z está dada por

$$f(x, y, z) = \begin{cases} \frac{1}{3}(2x + 3y + z) & \text{para } 0 < x < 1, 0 < y < 1, 0 < z < 1 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

encuentre

(a) $P(X = \frac{1}{2}, Y = \frac{1}{2}, Z = \frac{1}{2})$;

(b) $P(X < \frac{1}{2}, Y < \frac{1}{2}, Z < \frac{1}{2})$.

APLICACIONES

3.83 Suponga que tiramos un par de dados balanceados y X es el número de dados que salen 1, y Y es el número de dados que salen 4, 5 o 6.

(a) Dibuje un diagrama como el de la figura 3.1 que muestre los valores de X y Y asociados con cada uno de los 36 puntos igualmente probables del espacio muestral.

(b) Construya una tabla que muestre los valores de la distribución de probabilidad conjunta de X y Y.

3.84 Se seleccionan al azar dos libros de texto de un anaquel que contiene tres textos de estadística, dos textos de matemáticas y tres textos de física. Si X es el número de textos de estadística y Y el número de textos de matemáticas realmente escogidos, construya una tabla que muestre los valores de la distribución de probabilidad conjunta de X y Y.

3.85 Si X es el número de caras y Y es el número de caras menos el número de cruces obtenidas en tres lanzamientos de una moneda balanceada, elabore una tabla que muestre los valores de la distribución de probabilidad conjunta de X y Y.

3.86 Un tirador certero apunta a un blanco circular con radio 1. Si dibujamos un sistema de coordenadas rectangular con su origen en el centro del blanco, las coordenadas del punto de impacto, (X, Y), son variables aleatorias que tienen la densidad de probabilidad conjunta

$$f(x, y) = \begin{cases} \dfrac{1}{\pi} & \text{para } 0 < x^2 + y^2 < 1 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

Encuentre

(a) $P[(X, Y) \in A]$, donde A es el sector del círculo en el primer cuadrante limitado por las líneas $y = 0$ y $y = x$;

(b) $P[(X, Y) \in B]$, donde $B = \{(x, y) | 0 < x^2 + y^2 < \frac{1}{2}\}$.

3.87 Una cierta universidad aplica exámenes de aptitud a todos los alumnos de primer ingreso en ciencias y humanidades. Si X y Y son, respectivamente, las proporciones de respuestas correctas que un estudiante obtiene en las pruebas de las dos materias, la distribución de probabilidad conjunta de estas variables aleatorias se puede aproximar con la densidad de probabilidad conjunta

$$f(x, y) = \begin{cases} \dfrac{2}{5}(2x + 3y) & \text{para } 0 < x < 1, 0 < y < 1 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

¿Cuáles son las probabilidades de que un estudiante obtenga

(a) menos de 0.40 en ambas pruebas;

(b) más de 0.80 en la prueba de ciencias y menos de 0.50 en la prueba de humanidades?

3.88 Suponga que P, el precio de cierta mercancía (en dólares), y S, sus ventas totales (en 10,000 unidades), son variables aleatorias cuya distribución de probabilidad conjunta se puede aproximar bastante con la densidad de probabilidad conjunta

$$f(p, s) = \begin{cases} 5pe^{-ps} & \text{para } 0.20 < p < 0.40, s > 0 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

Encuentre las probabilidades que

(a) el precio será menos que 30 centavos y las ventas excederán 20,000 unidades;

(b) el precio estará entre 25 y 30 centavos y las ventas serán de menos de 10,000 unidades.

3.6 DISTRIBUCIONES MARGINALES

Para introducir el concepto de una **distribución marginal**, consideremos el siguiente ejemplo.

EJEMPLO 3.20

En el ejemplo 3.12 derivamos la distribución de probabilidad conjunta de dos variables aleatorias X y Y, el número de cápsulas de aspirinas y el número de cápsulas de sedante incluidas entre las dos cápsulas que se sacaron al azar de un frasco que contiene tres aspirinas, dos sedantes y cuatro cápsulas laxantes. Encuentre la distribución de probabilidad de X sola y de Y sola.

Solución

En la siguiente tabla se muestran los resultados del ejemplo 3.12, junto con los **totales marginales**, esto es, los totales de las respectivas hileras y columnas

$$
\begin{array}{cc|ccc|c}
 & & & x & & \\
 & & 0 & 1 & 2 & \\
\hline
 & 0 & \frac{1}{6} & \frac{1}{3} & \frac{1}{12} & \frac{7}{12} \\
y & 1 & \frac{2}{9} & \frac{1}{6} & & \frac{7}{18} \\
 & 2 & \frac{1}{36} & & & \frac{1}{36} \\
\hline
 & & \frac{5}{12} & \frac{1}{2} & \frac{1}{12} &
\end{array}
$$

Los totales de columna son las probabilidades de que X asumirá los valores 0, 1, y 2. En otras palabras, son los valores

$$g(x) = \sum_{y=0}^{2} f(x, y) \qquad \text{para } x = 0, 1, 2$$

de la distribución de probabilidad de X. Por la misma razón, los totales de las hileras son los valores

$$h(y) = \sum_{x=0}^{2} f(x, y) \qquad \text{para } y = 0, 1, 2$$

de la distribución de probabilidad de Y. ▲

Así llegamos a la siguiente definición.

DEFINICIÓN **3.10** Si X y Y son variables aleatorias discretas y $f(x, y)$ es el valor de la distribución de probabilidad conjunta en (x, y), la función dada por

$$g(x) = \sum_{y} f(x, y)$$

para cada x dentro del intervalo X es llamada la **distribución marginal** de X. Correspondientemente, la función dada por

$$h(y) = \sum_{x} f(x, y)$$

para cada y es llamada la **distribución marginal** de Y.

Cuando X y Y son variables aleatorias continuas, las distribuciones de probabilidad se reemplazan con densidades de probabilidad, las sumas se reemplazan con integrales y obtenemos

DEFINICIÓN 3.11 Si X y Y son variables aleatorias continuas y $f(x, y)$ es el valor de su densidad de probabilidad conjunta en (x, y), la función dada por

$$g(x) = \int_{-\infty}^{\infty} f(x, y)\, dy \qquad \text{para } -\infty < x < \infty$$

es llamada la **densidad marginal** de X. Correspondientemente, la función dada por

$$h(y) = \int_{-\infty}^{\infty} f(x, y)\, dx \qquad \text{para } -\infty < y < \infty$$

es llamada la **densidad marginal** de Y.

EJEMPLO 3.21

Dada la densidad de probabilidad conjunta

$$f(x, y) = \begin{cases} \dfrac{2}{3}(x + 2y) & \text{para } 0 < x < 1, 0 < y < 1 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

encuentre la diversidad marginal de X y Y.

Solución

Al efectuar las integraciones necesarias, obtenemos

$$g(x) = \int_{-\infty}^{\infty} f(x, y)\, dy = \int_{0}^{1} \frac{2}{3}(x + 2y)\, dy = \frac{2}{3}(x + 1)$$

para $0 < x < 1$ y $g(x) = 0$ en cualquier otra parte. De igual manera,

$$h(y) = \int_{-\infty}^{\infty} f(x, y)\, dx = \int_{0}^{1} \frac{2}{3}(x + 2y)\, dx = \frac{1}{3}(1 + 4y)$$

para $0 < y < 1$ y $h(y) = 0$ en cualquier otra parte. ▲

Cuando estamos trabajando con más de dos variables aleatorias, podemos hablar no sólo de las distribuciones marginales de las variables aleatorias individuales, sino también de las **distribuciones marginales conjuntas** de muchas de las variables aleatorias. Si la distribución de probabilidad conjunta de las variables aleatorias discretas X_1, $X_2, \ldots$ y X_n tiene los valores $f(x_1, x_2, \ldots, x_n)$, la distribución marginal de X_1 sola está dada por

$$g(x_1) = \sum_{x_2} \cdots \sum_{x_n} f(x_1, x_2, \ldots, x_n)$$

para todos los valores dentro del intervalo de X_1, la distribución marginal conjunta de X_1, X_2 y X_3 está dada por

$$m(x_1, x_2, x_3) = \sum_{x_4} \cdots \sum_{x_n} f(x_1, x_2, \ldots, x_n)$$

para todos los valores dentro del rango de X_1, X_2 y X_3, y otras distribuciones marginales se pueden definir en la misma forma. Para el caso continuo, las distribuciones de probabilidad se reemplazan con densidades de probabilidad, las sumas se reemplazan con integrales, y si la densidad de probabilidad conjunta de las variables aleatorias continuas X_1, $X_2, \ldots$ y X_n tiene los valores $f(x_1, x_2, \ldots, x_n)$, la densidad marginal de X_2 sola está dada por

$$h(x_2) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, x_2, \ldots, x_n)\, dx_1\, dx_3 \cdots dx_n$$

para $-\infty < x_2 < \infty$, la distribución marginal conjunta de X_1 y X_n están dadas por

$$\varphi(x_1, x_n) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, x_2, \ldots, x_n)\, dx_2\, dx_3 \cdots dx_{n-1}$$

para $-\infty < x_1 < \infty$ y $-\infty < x_n < \infty$, y así sucesivamente.

EJEMPLO 3.22

Consideremos otra vez la densidad de probabilidad trivariada del ejemplo 3.19,

$$f(x_1, x_2, x_3) = \begin{cases} (x_1 + x_2)e^{-x_3} & \text{para } 0 < x_1 < 1,\ 0 < x_2 < 1,\ x_3 > 0 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

encuentre la densidad marginal conjunta de X_1 y X_3 y la densidad marginal de X_1 sola.

Solución

Al efectuar la integración necesaria, encontramos que la densidad marginal conjunta de X_1 y X_3 está dada por

$$m(x_1, x_3) = \int_0^1 (x_1 + x_2)e^{-x_3}\, dx_2 = \left(x_1 + \frac{1}{2}\right)e^{-x_3}$$

para $0 < x_1 < 1$ y $x_3 > 0$ y $m(x_1, x_3) = 0$ en cualquier otra parte. Usando este resultado, encontramos que la densidad marginal de X_1 sola está dada por

$$g(x_1) = \int_0^{\infty} \int_0^1 f(x_1, x_2, x_3)\, dx_2\, dx_3 = \int_0^{\infty} m(x_1, x_3)\, dx_3$$

$$= \int_0^{\infty} \left(x_1 + \frac{1}{2}\right)e^{-x_3}\, dx_3 = x_1 + \frac{1}{2}$$

para $0 < x_1 < 1$ y $g(x_1) = 0$ en cualquier otra parte. ▲

Correspondiendo a las diversas densidades y distribuciones marginales y marginales conjuntas que hemos introducido en esta sección, también podemos definir **funciones de distribución marginales** y **marginales conjuntas**. Algunos problemas relacionados con tales funciones de distribución se dejan al lector en los ejercicios 3.92, 3.99 y 3.100.

3.7 DISTRIBUCIONES CONDICIONALES

En el capítulo 2 definimos la probabilidad condicional del evento A, dado el evento B, como

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

siempre y cuando $P(B) \neq 0$. Supongamos ahora que A y B son los eventos $X = x$ y $Y = y$ de modo que podemos escribir

$$P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}$$

$$= \frac{f(x, y)}{h(y)}$$

siempre y cuando $P(Y = y) = h(y) \neq 0$, donde $f(x, y)$ es el valor de la distribución de probabilidad conjunta de X y Y en (x, y), y $h(y)$ es el valor de la distribución marginal de Y en y. Denotemos la probabilidad condicional con $f(x|y)$ para indicar que x es una variable y y está fija, hagamos ahora la siguiente definición.

DEFINICIÓN **3.12** Si $f(x, y)$ es el valor de la distribución de probabilidad conjunta de las variables aleatorias discretas X y Y en (x, y), y $h(y)$ es el valor de la distribución marginal de Y en y, la función dada por

$$f(x|y) = \frac{f(x, y)}{h(y)} \qquad h(y) \neq 0$$

para cada x dentro del intervalo de X, se llama la **distribución condicional de** X dado $Y = y$. En forma correspondiente, si $g(x)$ es el valor de la distribución marginal de X en x, la función dada por

$$w(y|x) = \frac{f(x, y)}{g(x)} \qquad g(x) \neq 0$$

para cada y dentro del intervalo de Y, se llama la **distribución condicional** de Y dado $X = x$.

EJEMPLO 3.23

Con respecto a los ejemplos 3.12 y 3.20, encuentre la distribución condicional de X dado $Y = 1$.

Solución

Al sustituir los valores apropiados de la tabla en la página 116, obtenemos

$$f(0|1) = \frac{\frac{2}{9}}{\frac{7}{18}} = \frac{4}{7}$$

$$f(1|1) = \frac{\frac{1}{6}}{\frac{7}{18}} = \frac{3}{7}$$

$$f(2|1) = \frac{0}{\frac{7}{18}} = 0 \qquad \blacktriangle$$

Donde X y Y son variables aleatorias continuas, las distribuciones de probabilidad se reemplazan con densidades de probabilidad, y obtenemos

DEFINICIÓN **3.13** Si $f(x, y)$ es el valor de la densidad conjunta de las variables aleatorias continuas X y Y en (x, y), y $h(y)$ es el valor de la densidad marginal de Y en y, la función dada por

$$f(x|y) = \frac{f(x, y)}{h(y)} \qquad h(y) \neq 0$$

para $-\infty < x < \infty$, se llama **densidad condicional** de X dado $Y = y$. De manera correspondiente, si $g(x)$ es el valor de la densidad marginal de X en x, la función dada por

$$w(y|x) = \frac{f(x, y)}{g(x)} \qquad g(x) \neq 0$$

para $-\infty < y < \infty$, se llama densidad **condicional** de Y dado $X = x$.

EJEMPLO 3.24

Con respecto al ejemplo 3.21, encuentre la densidad condicional de X dado $Y = y$ y úsela para evaluar $P(X \leq \frac{1}{2}|Y = \frac{1}{2})$.

Solución

Al valernos de los resultados obtenidos en la página 117, tenemos

$$f(x|y) = \frac{f(x, y)}{h(y)} = \frac{\frac{2}{3}(x + 2y)}{\frac{1}{3}(1 + 4y)}$$

$$= \frac{2x + 4y}{1 + 4y}$$

para $0 < x < 1$ y $f(x|y) = 0$ en cualquier otra parte. Ahora

$$f\left(x\middle|\frac{1}{2}\right) = \frac{2x + 4 \cdot \frac{1}{2}}{1 + 4 \cdot \frac{1}{2}}$$

$$= \frac{2x + 2}{3}$$

y podemos escribir

$$P\left(X \leq \frac{1}{2}\middle|Y = \frac{1}{2}\right) = \int_0^{\frac{1}{2}} \frac{2x + 2}{3}\, dx = \frac{5}{12}$$

Es interesante observar que en la figura 3.12 esta probabilidad está dada por la razón del área del trapezoide $ABCD$ al área del trapezoide $AEFD$. ▲

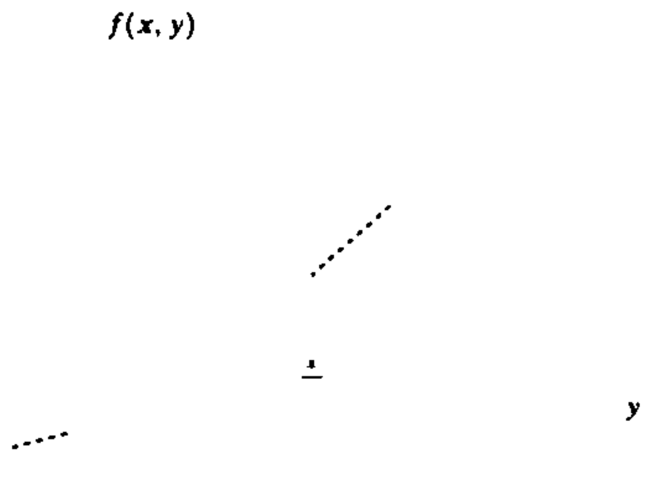

Figura 3.12 Diagrama para el ejemplo 3.24.

EJEMPLO 3.25

Dada la densidad de probabilidad conjunta

$$f(x, y) = \begin{cases} 4xy & \text{para } 0 < x < 1, 0 < y < 1 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

encuentre las densidades marginales de X y Y y la densidad condicional de X dado $Y = y$.

Solución

Al efectuar las integraciones necesarias, obtenemos

$$g(x) = \int_{-\infty}^{\infty} f(x, y)\, dy = \int_0^1 4xy\, dy$$

$$= 2xy^2 \Big|_{y=0}^{y=1} = 2x$$

para $0 < x < 1$, y $g(x) = 0$ en cualquier otra parte; también

$$h(y) = \int_{-\infty}^{\infty} f(x, y)\, dx = \int_0^1 4xy\, dx$$

$$= 2x^2y \Big|_{x=0}^{x=1} = 2y$$

para $0 < y < 1$, y $h(y) = 0$ en cualquier otra parte. Entonces, al sustituir en la fórmula para la densidad condicional, obtenemos

$$f(x|y) = \frac{f(x, y)}{h(y)} = \frac{4xy}{2y} = 2x$$

para $0 < x < 1$, y $f(x|y) = 0$ en cualquier otra parte. ▲

Cuando estamos trabajando con más de dos variables aleatorias, ya sea continuas o discretas, podemos considerar varias clases diferentes de distribuciones o densidades condicionales. Por ejemplo, si $f(x_1, x_2, x_3, x_4)$ es el valor de la distribución conjunta de las variables aleatorias discretas X_1, X_2, X_3 y X_4 en (x_1, x_2, x_3, x_4), podemos escribir

$$p(x_3|x_1, x_2, x_4) = \frac{f(x_1, x_2, x_3, x_4)}{g(x_1, x_2, x_4)} \qquad g(x_1, x_2, x_4) \neq 0$$

para el valor de la distribución condicional de X_3 en x_3 dado $X_1 = x_1$, $X_2 = x_2$ y $X_4 = x_4$, donde $g(x_1, x_2, x_4)$ es el valor de la distribución marginal conjunta de X_1, X_2 y X_4 en (x_1, x_2, x_4). También podemos escribir

$$q(x_2, x_4|x_1, x_3) = \frac{f(x_1, x_2, x_3, x_4)}{m(x_1, x_3)} \qquad m(x_1, x_3) \neq 0$$

para el valor de la **distribución condicional conjunta** de X_2 y X_4 en (x_2, x_4) dado $X_1 = x_1$ y $X_3 = x_3$, o

$$r(x_2, x_3, x_4|x_1) = \frac{f(x_1, x_2, x_3, x_4)}{b(x_1)} \qquad b(x_1) \neq 0$$

para el valor de la distribución condicional conjunta de X_2, X_3 y X_4 en (x_2, x_3, x_4) dado $X_1 = x_1$.

Cuando trabajamos con dos o más variables aleatorias, son generalmente de gran importancia las preguntas de **independencia**. En el ejemplo 3.25 vemos que $f(x|y) = 2x$ no depende del valor dado $Y = y$, pero éste claramente no es el caso en el ejemplo 3.24, donde $f(x|y) = \frac{2x + 4y}{1 + 4y}$. Siempre que los valores de la distribución condicional de X dado $Y = y$ no dependan de y, se sigue que $f(x|y) = g(x)$, y por tanto las fórmulas de las definiciones 3.12 y 3.13 nos dan

$$f(x, y) = f(x|y) \cdot h(y) = g(x) \cdot h(y)$$

Esto es, los valores de distribución conjunta están dados por los productos de los valores correspondientes de las dos distribuciones marginales. Al generalizar a partir de esta observación, hagamos ahora la siguiente definición.

> **DEFINICIÓN 3.14** Si $f(x_1, x_2, \ldots, x_n)$ es el valor de la distribución de probabilidad conjunta de las n variables aleatorias discretas $X_1, X_2, \ldots, X_n$ en $(x_1, x_2, \ldots, x_n)$, y $f_i(x_i)$ es el valor de la distribución marginal de X_i en x_i para $i = 1, 2, \ldots, n$, entonces las n variables aleatorias son **independientes** si y sólo si
>
> $$f(x_1, x_2, \ldots, x_n) = f_1(x_1) \cdot f_2(x_2) \cdot \ldots \cdot f_n(x_n)$$
>
> para toda $(x_1, x_2, \ldots, x_n)$ dentro de su intervalo.

Para dar la definición correspondiente para variables aleatorias continuas, simplemente sustituimos la palabra *densidad* por la palabra *distribución*.

Con esta definición de independencia, se puede verificar fácilmente que las tres variables aleatorias del ejemplo 3.22 no son independientes, pero que las variables aleatorias X_1 y X_3 X_1 y X_3 y también las dos variables aleatorias X_2 y X_3 son **independientes por parejas** (ver ejercicio 3.101).

Los ejemplos siguientes sirven para ilustrar el uso de la definición 3.14 para encontrar probabilidades relacionadas a varias variables aleatorias independientes.

EJEMPLO 3.26

Consideremos n lanzamientos independientes de una moneda balanceada, sea X_i el número de caras (0 o 1) obtenidas en el $i$ésimo lanzamiento para $i = 1, 2, \ldots, n$. Encuentre la distribución de probabilidad conjunta de estas n variables aleatorias

Solución

Puesto que cada una de las variables aleatorias X_i, para $i = 1, 2, \ldots, n$, tiene la distribución de probabilidad

$$f_i(x_i) = \frac{1}{2} \qquad \text{para } x_i = 0, 1$$

y las n variables aleatorias son independientes, su distribución de probabilidad conjunta está dada por

$$\begin{aligned} f(x_1, x_2, \ldots, x_n) &= f_1(x_1) \cdot f_2(x_2) \cdot \ldots \cdot f_n(x_n) \\ &= \frac{1}{2} \cdot \frac{1}{2} \cdot \ldots \cdot \frac{1}{2} = \left(\frac{1}{2}\right)^n \end{aligned}$$

donde $x_i = 0$ o 1 para $i = 1, 2, \ldots, n$. ▲

EJEMPLO 3.27

Dadas las variables aleatorias independientes X_1, X_2 y X_3 con densidades de probabilidad

$$f_1(x_1) = \begin{cases} e^{-x_1} & \text{para } x_1 > 0 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

$$f_2(x_2) = \begin{cases} 2e^{-2x_2} & \text{para } x_2 > 0 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

$$f_3(x_3) = \begin{cases} 3e^{-3x_3} & \text{para } x_3 > 0 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

encuentre la densidad de probabilidad conjunta y úsela para evaluar la probabilidad $P(X_1 + X_2 \leqq 1, X_3 > 1)$.

Solución

De acuerdo a la definición 3.14, los valores de la densidad de probabilidad conjunta están dadas por

$$\begin{aligned} f(x_1, x_2, x_3) &= f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_3) \\ &= e^{-x_1} \cdot 2e^{-2x_2} \cdot 3e^{-3x_3} \\ &= 6e^{-x_1 - 2x_2 - 3x_3} \end{aligned}$$

para $x_1 > 0$, $x_2 > 0$, $x_3 > 0$, y $f(x_1, x_2, x_3) = 0$ en cualquier otra parte. Así

$$\begin{aligned} P(X_1 + X_2 \leqq 1, X_3 > 1) &= \int_1^{\infty} \int_0^1 \int_0^{1-x_2} 6e^{-x_1 - 2x_2 - 3x_3}\, dx_1\, dx_2\, dx_3 \\ &= (1 - 2e^{-1} + e^{-2})e^{-3} \\ &= 0.020 \end{aligned}$$ ▲

EJERCICIOS

3.89 Dados los valores de la distribución de probabilidad conjunta de X y Y mostrados en la tabla

		$x = -1$	$x = 1$
	-1	$\frac{1}{8}$	$\frac{1}{2}$
y	0	0	$\frac{1}{4}$
	1	$\frac{1}{8}$	0

encuentre

(a) la distribución marginal de X;

(b) la distribución marginal de Y;

(c) la distribución condicional de X dado $Y = -1$.

3.90 Con respecto al ejercicio 3.56, encuentre

(a) la distribución marginal de X;

(b) la distribución marginal de Y;

(c) la distribución condicional de X dado $Y = 1$;

(d) la distribución condicional de Y dado $X = 0$.

3.91 Dada la distribución de probabilidad conjunta

$$f(x, y, z) = \frac{xyz}{108} \qquad \text{para } x = 1, 2, 3; \quad y = 1, 2, 3; \quad z = 1, 2$$

encuentre

(a) la distribución marginal conjunta de X y Y;

(b) la distribución marginal conjunta de X y Z;

(c) la distribución marginal de X;

(d) la distribución condicional de Z dado $X = 1$ y $Y = 2$;

(e) la distribución de probabilidad conjunta de Y y Z dado $X = 3$.

3.92 Con respecto al ejemplo 3.20, encuentre

(a) la **función de distribución marginal** de X, es decir, la función dada por $G(x) = P(X \leqq x)$ para $-\infty < x < \infty$;

(b) la **función de distribución condicional** de X dado $Y = 1$, es decir, la función dada por $F(x|1) = P(X \leqq x | Y = 1)$ para $-\infty < x < \infty$.

3.93 Verifique si X y Y son independientes si su distribución de probabilidad conjunta está dada por

(a) $f(x, y) = \frac{1}{4}$ para $x = -1$ y $y = -1$, $x = -1$ y $y = 1$, $x = 1$ y $y = -1$, y $x = 1$ y $y = 1$;

(b) $f(x, y) = \frac{1}{3}$ para $x = 0$ y $y = 0$, $x = 0$ y $y = 1$, y $x = 1$ y $y = 1$.

3.94 Si la densidad de probabilidad conjunta de X y Y está dada por

$$f(x, y) = \begin{cases} \frac{1}{4}(2x + y) & \text{para } 0 < x < 1, 0 < y < 2 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

encuentre

(a) la densidad marginal de X;

(b) la densidad condicional de Y dado $X = \frac{1}{4}$.

3.95 Con respecto al ejercicio 3.94, encuentre

(a) la densidad marginal de Y;

(b) la densidad condicional de X dado $Y = 1$.

3.96 Si la densidad de probabilidad conjunta de X y Y está dada por

$$f(x, y) = \begin{cases} 24y(1 - x - y) & \text{para } x > 0, y > 0, x + y < 1 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

encuentre

(a) la densidad marginal de X;

(b) la densidad marginal de Y.

Determine también si las dos variables aleatorias son independientes.

3.97 Con respecto al ejercicio 3.67, encuentre

(a) la densidad marginal de X;

(b) la densidad marginal de Y.

Determine también si las dos variables aleatorias son independientes.

3.98 Con respecto al ejemplo 3.22, encuentre

(a) la densidad condicional de X_2 dado $X_1 = \frac{1}{3}$ y $X_3 = 2$;

(b) la densidad de probabilidad conjunta de X_2 y X_3 dado $X_1 = \frac{1}{2}$.

3.99 Si $F(x, y)$ es el valor de la función de distribución conjunta de X y Y en (x, y), demuestre que la **función de distribución marginal** de X está dada por

$$G(x) = F(x, \infty) \qquad \text{para } -\infty < x < \infty$$

Use este resultado para encontrar la función de distribución marginal de X para las variables aleatorias del ejercicio 3.68.

3.100 Si $F(x_1, x_2, x_3)$ es el valor de la función de distribución conjunta de X_1, X_2 y X_3 en (x_1, x_2, x_3), demuestre que la **función de distribución marginal conjunta** X_1 y X_3 está dada por

$$M(x_1, x_3) = F(x_1, \infty, x_3) \qquad \text{para } -\infty < x_1 < \infty, -\infty < x_3 < \infty$$

y que la **función de distribución marginal** de X_1 está dada por

$$G(x_1) = F(x_1, \infty, \infty) \qquad \text{para } -\infty < x_1 < \infty$$

Con respecto al ejemplo 3.19, use estos resultados para encontrar

(a) la función de distribución marginal conjunta de X_1 y X_3;

(b) la función de distribución marginal de X_1.

3.101 Con referencia al ejemplo 3.22, verifique que las tres variables aleatorias X_1, X_2 y X_3 no son independientes, pero que las dos variables aleatorias X_1 y X_3 y también las dos variables aleatorias X_2 y X_3 son **independientes por parejas**.

3.102 Si las variables aleatorias independientes X y Y tienen las densidades marginales

$$f(x) = \begin{cases} \frac{1}{2} & \text{para } 0 < x < 2 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

$$\pi(y) = \begin{cases} \frac{1}{3} & \text{para } 0 < y < 3 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

encuentre

(a) la densidad de probabilidad conjunta de X y Y;

(b) el valor de $P(X^2 + Y^2 > 1)$.

APLICACIONES

3.103 Con respecto al ejercicio 3.84, encuentre

(a) la distribución marginal de X;

(b) la distribución condicional de Y dado $X = 0$.

3.104 Si se sacan dos cartas aleatoriamente (sin reemplazo) de una baraja común de 52 cartas de juego, Z es el número de ases obtenidos en la primera carta y W es el número total de ases obtenidos en las dos cartas, encuentre

(a) la distribución de probabilidad conjunta de Z y W;

(b) la distribución marginal de Z;

(c) la distribución condicional de W dado $Z = 1$.

3.105 Si X es la proporción de personas que responderán a una clase de solicitud por correo, Y es la proporción que responderán a otra clase de solicitud por correo, y la densidad de probabilidad conjunta de X y Y está dada por

$$f(x, y) = \begin{cases} \frac{2}{5}(x + 4y) & \text{para } 0 < x < 1, 0 < y < 1 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

encuentre las probabilidades de que

(a) al menos 30% responderá a la primera clase de solicitud por correo;

(b) cuando mucho 50% responderá a la segunda clase de solicitud por correo dado que ha habido un 20% de respuesta a la primera clase de solicitud por correo.

3.106 Con respecto al ejercicio 3.88, encuentre

(a) la densidad marginal de P;

(b) la densidad condicional de S dado $P = p$;

(c) la probabilidad de que las ventas serán menores que 30,000 unidades cuando $p = 25$ cents.

3.107 Si X es la cantidad de dinero (en dólares) que un agente de ventas gasta en gasolina durante un día y Y es la cantidad de dinero (en dólares) correspondiente que le reembolsan, la densidad conjunta de estas dos variables aleatorias está dada por

$$f(x, y) = \begin{cases} \dfrac{1}{25}\left(\dfrac{20 - x}{x}\right) & \text{para } 10 < x < 20, \dfrac{x}{2} < y < x \\ 0 & \text{en cualquier otra parte} \end{cases}$$

encuentre

(a) la densidad marginal de X;

(b) la densidad condicional de X dado $X = 12$;

(c) la probabilidad de que al agente de ventas se le reembolsarán al menos \$8 cuando gasta \$12.

3.108 Muestre que las dos variables aleatorias del ejercicio 3.87 no son independientes.

3.109 La vida útil (en horas) de cierta clase de tubo al vacío es una variable aleatoria que tiene la densidad de probabilidad

$$f(x) = \begin{cases} \dfrac{20{,}000}{(x + 100)^3} & \text{para } x > 0 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

Si tres de estos tubos funcionan independientemente, encuentre

(a) la densidad de probabilidad conjunta de X_1, X_2 y X_3, que representan las longitudes de sus vidas útiles;

(b) el valor de $P(X_1 < 100, X_2 < 100, X_3 \geqq 200)$.

REFERENCIAS

Tratamientos más avanzados o más detallados del material en este capítulo se pueden encontrar en

BRUNK, H. D., *An Introduction to Mathematical Statistics*, 3rd ed. Lexington, Mass.: Xerox College Publishing, 1975,

DEGROOT, M. H., *Probability and Statistics*, 2nd ed. Reading, Mass.: Addison-Wesley Publishing Company, Inc., 1986,

FRASER, D. A. S., *Probability and Statistics: Theory and Applications*. North Scituate, Mass.: Duxbury Press, 1976,

HOGG, R. V., and CRAIG, A. T., *Introduction to Mathematical Statistics*, 4th ed. Nueva York: Macmillan Publishing Co., Inc., 1978,

KENDALL, M. G., and STUART, A., *The Advanced Theory of Statistics*, Vol. 1, 4th ed. Nueva York: Macmillan Publishing Co., Inc., 1977,

KHAZANIE, R., *Basic Probability Theory and Applications*. Pacific Palisades, Calif.: Goodyear Publishing Company, Inc., 1976.

CAPÍTULO

4

Esperanza matemática

4.1 INTRODUCCIÓN

Originalmente, el concepto de **esperanza matemática** surgió en relación con los juegos de azar, y en su forma más simple es el producto de la cantidad que un jugador puede ganar y la probabilidad de que ganará. Por ejemplo, si tenemos uno de 10,000 boletos en una rifa cuyo premio principal es un viaje que vale \$4,800, nuestra esperanza matemática es $4{,}800 \cdot \frac{1}{10{,}000} = \$0.48$. Esta cantidad deberá interpretarse en el sentido de un promedio: en conjunto los 10,000 boletos pagarán \$4,800 o en promedio $\frac{\$4{,}800}{10{,}000} = \$0.48$ por boleto.

Si también hay un segundo premio que vale \$1,200 y un tercer premio con valor de \$400, podemos argumentar que en conjunto los 10,000 boletos pagan \$4,800 + \$1,200 + \$400 = \$6,400, o en promedio $\frac{\$6{,}400}{10{,}000} = \$0.64$ por boleto. Veamos esto en una forma diferente, podemos argumentar que si la rifa se repite muchas veces, perderíamos 99.97 por ciento de las veces (o con una probabilidad de 0.9997) y ganaríamos cada uno de los premios 0.01 por ciento de las veces (o con probabilidad de 0.0001). En promedio ganaríamos así

$$0(0.9997) + 4{,}800(0.0001) + 1{,}200(0.0001) + 400(0.0001) = \$0.64$$

que es la suma de los productos obtenidos al multiplicar cada cantidad por la probabilidad correspondiente.

4.2 EL VALOR ESPERADO DE UNA VARIABLE ALEATORIA

En la ilustración de la sección anterior, la cantidad que ganamos fue una variable aleatoria, y la esperanza matemática de esta variable aleatoria fue la suma de los productos obtenidos al multiplicar cada valor de la variable aleatoria por la probabilidad correspondiente. Nos referimos a la esperanza matemática de una variable aleatoria simplemente como su **valor esperado**, y extendemos la definición al caso continuo al reemplazar la operación de suma por la integración, así tenemos

DEFINICIÓN **4.1** Si X es una variable aleatoria discreta y $f(x)$ es el valor de su distribución de probabilidad en x, el **valor esperado** de X es

$$E(X) = \sum_x x \cdot f(x)$$

De manera correspondiente, si X es variable aleatoria continua y $f(x)$ es el valor de su densidad de probabilidad en x, el **valor esperado** de X es

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\, dx$$

En esta definición se supone, por supuesto, que la suma o la integral existe; de otra manera, la esperanza matemática está indefinida.

EJEMPLO 4.1

En un lote de 12 aparatos de televisión se incluye 2 con cables blancos. Si se escogen al azar tres de los aparatos para enviarse a un hotel, ¿cuántos de los aparatos con cables blancos puede el expedidor esperar enviar al hotel?

Solución

Puesto que x de los dos aparatos con cables blancos y $3 - x$ de los otros 10 aparatos se pueden escoger en $\binom{2}{x}\binom{10}{3-x}$ maneras, tres de los 12 aparatos se pueden escoger de $\binom{12}{3}$ maneras, y estas $\binom{12}{3}$ posibilidades son presuntamente equiprobables, encontramos que la distribución de probabilidad de X, el número de aparatos con cables blancos enviados al hotel, está dado por

$$f(x) = \frac{\binom{2}{x}\binom{10}{3-x}}{\binom{12}{3}} \qquad \text{para } x = 0, 1, 2$$

o, en forma tabular:

x	0	1	2
$f(x)$	$\frac{6}{11}$	$\frac{9}{22}$	$\frac{1}{22}$

Ahora,

$$E(X) = 0 \cdot \frac{6}{11} + 1 \cdot \frac{9}{22} + 2 \cdot \frac{1}{22} = \frac{1}{2}$$

y puesto que no es posible enviar medio aparato, está claro que el término "esperar" no se usa en su sentido familiar. Ciertamente, deberá interpretarse como un promedio tocante a envíos repetidos hechos en las condiciones dadas. ▲

EJEMPLO 4.2

Ciertas medidas cifradas del diámetro de separación de las roscas de un adaptador tienen la densidad de probabilidad

$$f(x) = \begin{cases} \dfrac{4}{\pi(1 + x^2)} & \text{para } 0 < x < 1 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

Encuentre el valor esperado de esta variable aleatoria

Solución

Al usar la definición 4.1, tenemos

$$\begin{aligned} E(X) &= \int_0^1 x \cdot \frac{4}{\pi(1 + x^2)}\, dx \\ &= \frac{4}{\pi} \int_0^1 \frac{x}{1 + x^2}\, dx \\ &= \frac{\ln 4}{\pi} = 0.4413 \end{aligned}$$ ▲

Hay muchos problemas donde nos interesa no sólo el valor esperado de una variable aleatoria X, sino también los valores esperados de variables aleatorias relacionadas con X. Así, nos podría interesar la variable aleatoria Y, cuyos valores están relacionados con los de X por medio de la ecuación $y = g(x)$; para simplificar nuestra notación, denotamos esta variable aleatoria con $g(X)$. Por ejemplo, $g(X)$ podría ser X^3 así que cuando X asume el valor de 2, $g(X)$ asume el valor $2^3 = 8$. Si queremos encontrar el valor esperado de una variable aleatoria $g(X)$, tal, podríamos primero determinar su distribución o densidad de probabilidad (mediante uno de los métodos que se examinarán en el capítulo 7) y entonces usamos la definición 4.1, pero generalmente es más fácil y más directo usar el siguiente teorema

TEOREMA 4.1 Si X es una variable aleatoria discreta y $f(x)$ es el valor de su distribución de probabilidad en x, el valor esperado de $g(X)$ está dado por

$$E[g(X)] = \sum_x g(x) \cdot f(x)$$

De forma correspondiente, si X es una variable aleatoria continua y $f(x)$ es el valor de su densidad de probabilidad en x, el valor esperado de $g(X)$ está dado por

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) \cdot f(x)\, dx$$

Demostración. Puesto que una demostración más general rebasa el alcance de este texto, demostraremos aquí este teorema sólo para el caso donde X es discreta y tiene un intervalo finito. Puesto que $y = g(x)$ no necesariamente define una correspondencia unívoca, supongamos que $g(x)$ asume el valor g_i cuando x asume los valores $x_{i1}, x_{i2}, \ldots, x_{in_i}$. Entonces, la probabilidad que $g(X)$ asumirá el valor g_i es

$$P[g(X) = g_i] = \sum_{j=1}^{n_i} f(x_{ij})$$

y si $g(x)$ asume los valores $g_1, g_2, \ldots, g_m$, se sigue que

$$\begin{aligned}
E[g(X)] &= \sum_{i=1}^{m} g_i \cdot P[g(X) = g_i] \\
&= \sum_{i=1}^{m} g_i \cdot \sum_{j=1}^{n_i} f(x_{ij}) \\
&= \sum_{i=1}^{m} \sum_{j=1}^{n_i} g_i \cdot f(x_{ij}) \\
&= \sum_x g(x) \cdot f(x)
\end{aligned}$$

donde la suma se extiende sobre todos los valores de X. ▼

EJEMPLO 4.3

Si X es el número de puntos tirados con un dado balanceado, encuentre el valor esperado de $g(X) = 2X^2 + 1$.

Solución

Puesto que cada resultado posible tiene la probabilidad $\frac{1}{6}$, obtenemos

$$E[g(X)] = \sum_{x=1}^{6} (2x^2 + 1) \cdot \frac{1}{6}$$

$$= (2 \cdot 1^2 + 1) \cdot \frac{1}{6} + \cdots + (2 \cdot 6^2 + 1) \cdot \frac{1}{6}$$

$$= \frac{94}{3} \qquad \blacktriangle$$

EJEMPLO 4.4

Si X tiene la densidad de probabilidad

$$f(x) = \begin{cases} e^{-x} & \text{para } x > 0 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

encuentre el valor esperado de $g(X) = e^{3X/4}$.

Solución

De acuerdo al teorema 4.1, tenemos

$$E[e^{3X/4}] = \int_0^\infty e^{3x/4} \cdot e^{-x}\, dx$$

$$= \int_0^\infty e^{-x/4}\, dx$$

$$= 4 \qquad \blacktriangle$$

La determinación de las esperanzas matemáticas a menudo se puede simplificar al usar los siguientes teoremas, que nos permiten calcular los valores esperados a partir de otras esperanzas conocidas o fácilmente calculadas. Puesto que los pasos son esencialmente los mismos, algunas demostraciones se harán para el caso discreto o para el caso continuo; otras se dejan como ejercicios para el lector

TEOREMA 4.2 Si a y b son constantes, entonces

$$E(aX + b) = aE(X) + b$$

Demostración. Al usar el teorema 4.1 con $g(X) = aX + b$, obtenemos

$$\begin{aligned} E(aX + b) &= \int_{-\infty}^{\infty} (ax + b) \cdot f(x)\, dx \\ &= a \int_{-\infty}^{\infty} x \cdot f(x)\, dx + b \int_{-\infty}^{\infty} f(x)\, dx \\ &= aE(X) + b \qquad \blacktriangledown \end{aligned}$$

Si hacemos $b = 0$ o $a = 0$, se sigue del teorema 4.2 que

COROLARIO 1 Si a es una constante, entonces

$$E(aX) = aE(X)$$

COROLARIO 2 Si b es una constante, entonces

$$E(b) = b$$

Observe que si escribimos $E(b)$, la constante b se puede considerar como una variable aleatoria que siempre asume el valor b.

TEOREMA 4.3 Si $c_1, c_2, \ldots,$ y c_n son constantes, entonces

$$E\left[\sum_{i=1}^{n} c_i g_i(X)\right] = \sum_{i=1}^{n} c_i E[g_i(X)]$$

Demostración. De acuerdo al teorema 4.1 con $g(X) = \sum_{i=1}^{n} c_i g_i(X)$, obtenemos

$$\begin{aligned} E\left[\sum_{i=1}^{n} c_i g_i(X)\right] &= \sum_{x} \left[\sum_{i=1}^{n} c_i g_i(x)\right] f(x) \\ &= \sum_{i=1}^{n} \sum_{x} c_i g_i(x) f(x) \\ &= \sum_{i=1}^{n} c_i \sum_{x} g_i(x) f(x) \\ &= \sum_{i=1}^{n} c_i E[g_i(X)] \qquad \blacktriangledown \end{aligned}$$

EJEMPLO 4.5

Recurrimos al hecho de que

$$E(X^2) = (1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2)\cdot\frac{1}{6} = \frac{91}{6}$$

para la variable aleatoria del ejemplo 4.3, rehaga ese ejemplo.

Solución

$$E(2X^2 + 1) = 2E(X^2) + 1 = 2\cdot\frac{91}{6} + 1 = \frac{94}{3} \qquad \blacktriangle$$

EJEMPLO 4.6

Si la densidad de probabilidad de X está dada por

$$f(x) = \begin{cases} 2(1 - x) & \text{para } 0 < x < 1 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

(a) demuestre que

$$E(X^r) = \frac{2}{(r + 1)(r + 2)}$$

(b) y use este resultado para evaluar

$$E[(2X + 1)^2]$$

Solución

(a) $$E(X^r) = \int_0^1 x^r\cdot 2(1 - x)\,dx = 2\int_0^1 (x^r - x^{r+1})\,dx$$
$$= 2\left(\frac{1}{r + 1} - \frac{1}{r + 2}\right) = \frac{2}{(r + 1)(r + 2)}$$

(b) Puesto que $E[(2X + 1)^2] = 4E(X^2) + 4E(X) + 1$ y la sustitución de $r = 1$ y $r = 2$ en la fórmula anterior nos da $E(X) = \frac{2}{2\cdot 3} = \frac{1}{3}$ y $E(X^2) = \frac{2}{3\cdot 4} = \frac{1}{6}$, obtenemos

$$E[(2X + 1)^2] = 4\cdot\frac{1}{6} + 4\cdot\frac{1}{3} + 1 = 3 \qquad \blacktriangle$$

EJEMPLO 4.7

Demuestre que

$$E[(aX + b)^n] = \sum_{i=0}^{n}\binom{n}{i}a^{n-i}b^iE(X^{n-i})$$

Solución

Puesto que $(ax + b)^n = \sum_{i=0}^{n} \binom{n}{i} (ax)^{n-i} b^i$ de acuerdo al teorema 1.9, se sigue que

$$E[(aX + b)^n] = E\left[\sum_{i=0}^{n} \binom{n}{i} a^{n-i} b^i X^{n-i}\right]$$

$$= \sum_{i=0}^{n} \binom{n}{i} a^{n-i} b^i E(X^{n-i}) \qquad \blacktriangle$$

El concepto de una esperanza matemática se puede ampliar fácilmente a situaciones que implican más de una variable aleatoria. Por ejemplo, si Z es la variable aleatoria cuyos valores están relacionados con los de las dos variables aleatorias X y Y por medio de la ecuación $z = g(x, y)$, se puede demostrar que

TEOREMA 4.4 Si X y Y son variables aleatorias discretas y $f(x, y)$ es el valor de su distribución de probabilidad conjunta en (x, y), el valor esperado de $g(X, Y)$ es

$$E[g(X, Y)] = \sum_{x} \sum_{y} g(x, y) \cdot f(x, y)$$

De manera correspondiente, si X y Y son variables aleatorias continuas y $f(x, y)$ es el valor de su densidad de probabilidad conjunta en (x, y), el valor esperado de $g(X, Y)$ es

$$E[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y)\, dx\, dy$$

La generalización de este teorema a funciones de cualquier número finito de variables aleatorias es directa

EJEMPLO 4.8

Con respecto al ejemplo 3.12, encuentre el valor esperado de $g(X, Y) = X + Y$.

Solución

$$E(X + Y) = \sum_{x=0}^{2} \sum_{y=0}^{2} (x + y) \cdot f(x, y)$$

$$= (0 + 0) \cdot \frac{1}{6} + (0 + 1) \cdot \frac{2}{9} + (0 + 2) \cdot \frac{1}{36} + (1 + 0) \cdot \frac{1}{3}$$

$$+ (1 + 1) \cdot \frac{1}{6} + (2 + 0) \cdot \frac{1}{12}$$

$$= \frac{10}{9} \qquad \blacktriangle$$

EJEMPLO 4.9

Si la densidad de probabilidad conjunta de X y Y está dada por

$$f(x, y) = \begin{cases} \frac{2}{7}(x + 2y) & \text{para } 0 < x < 1,\ 1 < y < 2 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

encuentre el valor esperado de $g(X, Y) = X/Y^3$.

Solución

$$\begin{aligned} E(X/Y^3) &= \int_1^2 \int_0^1 \frac{2x(x + 2y)}{7y^3}\, dx\, dy \\ &= \frac{2}{7}\int_1^2 \left(\frac{1}{3y^3} + \frac{1}{y^2}\right) dy \\ &= \frac{15}{84} \qquad \blacktriangle \end{aligned}$$

Lo siguiente es otro teorema que tiene aplicaciones útiles en trabajo subsecuente. Es una generalización del teorema 4.3, y su demostración es paralela a la demostración de ese teorema.

TEOREMA 4.5 Si $c_1, c_2, \ldots,$ y c_n son constantes, entonces

$$E\left[\sum_{i=1}^{n} c_i g_i(X_1, X_2, \ldots, X_k)\right] = \sum_{i=1}^{n} c_i E[g_i(X_1, X_2, \ldots, X_k)]$$

EJERCICIOS

4.1 Para ilustrar la demostración del teorema 4.1, considere la variable aleatoria X, que asume los valores $-2, -1, 0, 1, 2$ y 3 con probabilidades $f(-2)$, $f(-1)$, $f(0)$, $f(1)$, $f(2)$ y $f(3)$. Si $g(X) = X^2$, encuentre

(a) g_1, g_2, g_3 y g_4, los cuatro valores posibles de $g(x)$;

(b) las probabilidades $P[g(X) = g_i]$ para $i = 1, 2, 3, 4$;

(c) $E[g(X)] = \sum_{i=1}^{4} g_i \cdot P[g(X) = g_i]$, y muestre que es igual a $\sum_x g(x) \cdot f(x)$.

4.2 Demuestre el teorema 4.2 para variables aleatorias discretas.

4.3 Demuestre el teorema 4.3 para variables aleatorias continuas.

4.4 Demuestre el teorema 4.5 para variables aleatorias discretas.

4.5 Dadas dos variables aleatorias continuas X y Y, use el teorema 4.4 para expresar $E(X)$ en términos de

(a) la densidad conjunta de X y Y;

(b) la densidad marginal de X.

4.6 Obtenga el valor esperado de la variable aleatoria discreta X que tiene la distribución de probabilidad

$$f(x) = \frac{|x-2|}{7} \qquad \text{para } x = -1, 0, 1, 3$$

4.7 Obtenga el valor esperado de la variable aleatoria Y cuya densidad de probabilidad está dada por

$$f(y) = \begin{cases} \frac{1}{8}(y+1) & \text{para } 2 < y < 4 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

4.8 Obtenga el valor esperado de la variable aleatoria X cuya densidad de probabilidad está dada por

$$f(x) = \begin{cases} x & \text{para } 0 < x < 1 \\ 2 - x & \text{para } 1 \leqq x < 2 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

4.9 (a) Si X asume los valores 0, 1, 2 y 3 con probabilidades $\frac{1}{125}$, $\frac{12}{125}$, $\frac{48}{125}$ y $\frac{64}{125}$, encuentre $E(X)$ y $E(X^2)$.

(b) Use los resultados del inciso (a) para determinar el valor de $E[(3X+2)^2]$.

4.10 (a) Si la densidad de probabilidad de X está dada por

$$f(x) = \begin{cases} \frac{1}{x(\ln 3)} & \text{para } 1 < x < 3 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

encuentre $E(X)$, $E(X^2)$ y $E(X^3)$.

(b) Use los resultados de la parte (a) para determinar $E(X^3 + 2X^2 - 3X + 1)$.

4.11 Si la densidad de probabilidad de X está dada por

$$f(x) = \begin{cases} \frac{x}{2} & \text{para } 0 < x \leqq 1 \\ \frac{1}{2} & \text{para } 1 < x \leqq 2 \\ \frac{3-x}{2} & \text{para } 2 < x < 3 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

encuentre los valores esperados de $g(X) = X^2 - 5X + 3$.

4.12 Con respecto al ejercicio 3.61, encuentre $E(2X - Y)$.

4.13 Con respecto al ejercicio 3.67, encuentre $E(X/Y)$.

4.14 Con respecto al ejercicio 3.76, encuentre el valor esperado de $U = X + Y + Z$.

4.15 Con respecto al ejercicio 3.82, encuentre el valor esperado de $W = X^2 - YZ$.

4.16 Si la distribución de probabilidad de X está dada por

$$f(x) = \left(\frac{1}{2}\right)^x \qquad \text{para } x = 1, 2, 3, \ldots$$

demuestre que $E(2^X)$ no existe. Ésta es la famosa **paradoja de Petersburgo**, según la cual la esperanza de un jugador es infinita (no existe) si recibirá 2^x dólares cuando, en una serie de lanzamientos de una moneda balanceada, la primera cara aparece en el $x$ésimo tiro.

APLICACIONES

4.17 La probabilidad de que la Sra. Vélez venderá una propiedad con una ganancia de \$3,000 es $\frac{3}{20}$, la probabilidad de que la venderá con una ganancia de \$1,500 es $\frac{7}{20}$, la probabilidad de que salga a mano es $\frac{7}{20}$, y la probabilidad de que perderá \$1,500 es $\frac{3}{20}$. ¿Cuál es su ganancia esperada?

4.18 Un juego de azar se considera **justo**, o **equitativo**, si la esperanza de cada jugador es igual a cero. Si alguien nos paga \$10 cada vez que tiramos un 3 o un 4 con un dado balanceado, para hacer el juego equitativo, ¿cuánto debemos pagar a esa persona cuando lanzamos un 1, un 2, un 5 o un 6?

4.19 El gerente de una pastelería sabe que el número de pasteles de chocolate que puede vender en un día dado es una variable aleatoria que tiene la distribución de probabilidad $f(x) = \frac{1}{6}$ para $x = 0, 1, 2, 3, 4$ y 5. También sabe que hay una ganancia de \$1 por cada pastel que él venda y una pérdida (causada por la descomposición) de \$0.40 por cada pastel que no vende. Supongamos que cada pastel puede venderse sólo en el día en que se hace, encuentre la ganancia esperada del pastelero para un día en que hornee

(a) uno de los pasteles;

(b) dos de los pasteles;

(c) tres de los pasteles;

(d) cuatro de los pasteles;

(e) cinco de los pasteles.

¿Cuántos debe hornear para maximizar su ganancia esperada?

4.20 Si la ganancia de una contratista en un trabajo de construcción se puede considerar como una variable aleatoria continua que tiene la densidad de probabilidad

$$f(x) = \begin{cases} \frac{1}{18}(x+1) & \text{para } -1 < x < 5 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

donde las unidades están en \$1,000, ¿cuál es su ganancia esperada?

4.21 Con respecto al ejercicio 3.53, ¿qué desgaste de uno de los neumáticos puede esperar obtener el dueño de un auto?

4.22 Con respecto al ejercicio 3.54, ¿cuál es el consumo de agua esperado en la ciudad para cualquier día dado?

4.23 Con respecto al ejercicio 3.88, encuentre $E(PS)$, los ingresos esperados de la mercancía.

4.24 El señor Arias y la señorita Sánchez están apostando en lanzamientos repetidos de una moneda. Al principio del juego el señor Arias tiene a dólares y la señorita Sánchez tiene b dólares, en cada lanzamiento el perdedor paga al ganador un dólar, y el juego continúa hasta que uno de los dos jugadores está "arruinado." Recurra al hecho de que en un juego equitativo la esperanza matemática de cada jugador es cero, encuentre la probabilidad de que el señor Arias ganará a la señorita Sánchez sus b dólares antes de perder sus a dólares.

4.3 MOMENTOS

En estadística, las esperanzas matemáticas definidas aquí y en la definición 4.4, llamadas los **momentos** de la distribución de una variable aleatoria o simplemente los **momentos** de una variable aleatoria, son de especial importancia.

DEFINICIÓN **4.2** El $r$ésimo momento alrededor del origen de una variable aleatoria X, denotado por μ'_r, es el valor esperado de X^r; simbólicamente,

$$\mu'_r = E(X^r) = \sum_x x^r \cdot f(x)$$

para $r = 0, 1, 2, \ldots$ donde X es discreta, y

$$\mu'_r = E(X^r) = \int_{-\infty}^{\infty} x^r \cdot f(x)\, dx$$

donde X es continua.

Es de interés señalar que el término "momento" viene del campo de la física: si las cantidades $f(x)$ en el caso discreto fueran puntos de masa que actúan perpendicularmente al eje x a distancias x del origen, μ'_1 sería la coordenada x del centro de gravedad, esto es, el primer momento dividido por $\sum f(x) = 1$, y μ'_2 sería el momento de inercia. Esto también explica por qué los momentos μ'_r se llaman momentos alrededor del origen: en analogía a la física, la longitud del brazo de palanca en cada caso es la distancia desde el origen. La analogía también se aplica en el caso continuo, donde μ'_1 y μ'_2 podrían ser la coordenada x del centro de gravedad y el momento de inercia de una varilla de densidad variable.

Cuando $r = 0$, tenemos $\mu'_0 = E(X^0) = E(1) = 1$ el corolario 2 del teorema 4.2 y este resultado está, como debiera ser, de acuerdo con los teoremas 3.1 y 3.5. Cuando

$r = 1$, tenemos $\mu_1' = E(X)$, que es justamente el valor esperado de la variable aleatoria X, y en vista de su importancia en la estadística le damos un símbolo especial y un nombre especial

> **DEFINICIÓN 4.3** μ_1' se llama la **media** de la distribución de X, o simplemente la media de X, y se denota con μ.

Los momentos especiales que definiremos a continuación son de importancia en la estadística porque sirven para describir la forma de la distribución de una variable aleatoria, esto es, la forma de la gráfica de su distribución o densidad de probabilidad.

> **DEFINICIÓN 4.4** El $r$ésimo **momento alrededor del origen** de una variable aleatoria X, denotado por μ_r, es el valor esperado de $(X - \mu)^r$; simbólicamente,
>
> $$\mu_r = E\left[(X - \mu)^r\right] = \sum_x (x - \mu)^r \cdot f(x)$$
>
> para $r = 0, 1, 2, \ldots$ cuando X es discreta, y
>
> $$\mu_r = E\left[(X - \mu)^r\right] = \int_{-\infty}^{\infty} (x - \mu)^r \cdot f(x)\, dx$$
>
> cuando X es continua

Advierta que $\mu_0 = 1$ y $\mu_1 = 0$ para cualquier variable aleatoria para la cual μ exista (véase el ejercicio 4.25).

El segundo momento alrededor de la media es de especial importancia en estadística porque indica la amplitud o dispersión de la distribución de una variable aleatoria; así, se le da un símbolo especial y un nombre especial.

> **DEFINICIÓN 4.5** μ_2 se llama la **varianza** de la distribución de X, o simplemente la **varianza** de X, y se denota por σ^2, var(X) o $V(X)$; σ, la raíz cuadrada positiva de la varianza, se llama la **desviación estándar**.

La figura 4.1 muestra cómo la varianza refleja la amplitud o dispersión de la distribución de una variable aleatoria. Mostramos aquí los histogramas de las distribuciones de probabilidad de cuatro variables aleatorias con la misma media $\mu = 5$ pero con varianzas iguales a 5.26, 3.18, 1.66 y 0.88. Como se puede ver, un valor pequeño de σ^2 sugiere que es probable que obtengamos un valor cercano a la media, y un valor grande de σ^2 sugiere que hay una mayor probabilidad de sacar un valor que no está cercano a la media. Esto se examinará más ampliamente en la sección 4.4. Un breve examen de có-

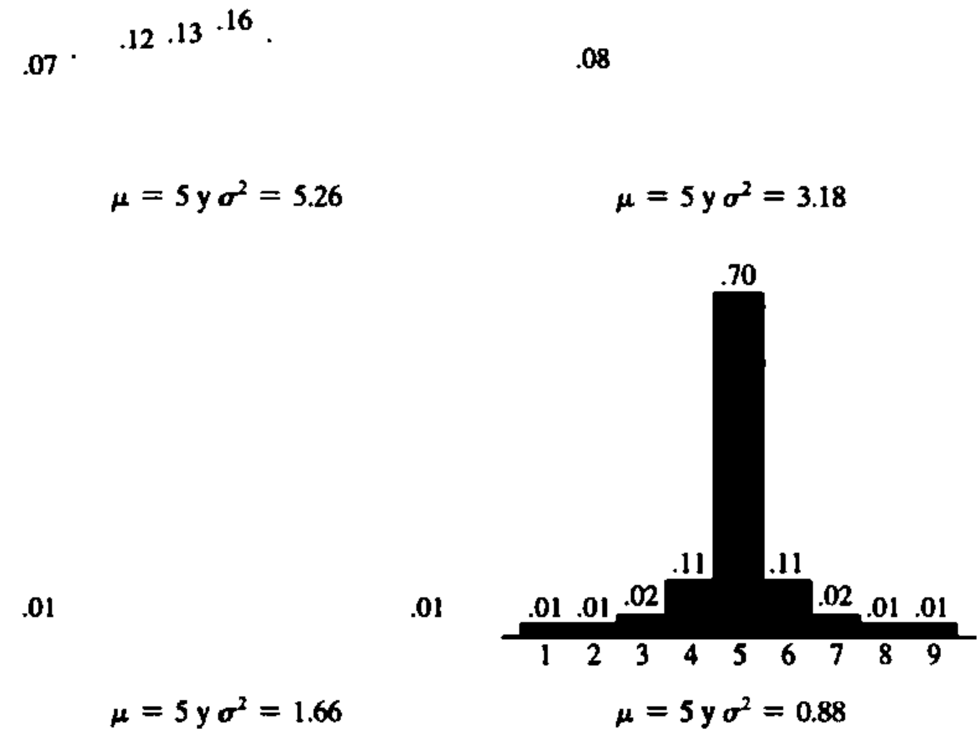

Figura 4.1 Distribuciones con diferentes dispersiones.

mo μ_3, momento alrededor de la media, describe la **simetría** o **asimetría** (falta de simetría) de una distribución se da en el ejercicio 4.34.

En muchos casos, los momentos alrededor de la media se obtienen al calcular primero los momentos alrededor del origen y entonces expresar el μ_r en términos del μ'_r. Con este fin, se le pedirá al lector que verifique una fórmula general en el ejercicio 4.33. Aquí, simplemente derivaremos la siguiente fórmula de cálculo para σ^2.

TEOREMA 4.6

$$\sigma^2 = \mu'_2 - \mu^2$$

Demostración

$$\begin{aligned}\sigma^2 &= E[(X-\mu)^2] \\ &= E(X^2 - 2\mu X + \mu^2) \\ &= E(X^2) - 2\mu E(X) + E(\mu^2) \\ &= E(X^2) - 2\mu\cdot\mu + \mu^2 \\ &= \mu'_2 - \mu^2 \qquad \blacktriangledown\end{aligned}$$

EJEMPLO 4.10

Use el teorema 4.6 para calcular la varianza de X, que representa el número de puntos lanzados con un dado balanceado.

Solución

Primero calculemos

$$\mu = E(X) = 1\cdot\frac{1}{6} + 2\cdot\frac{1}{6} + 3\cdot\frac{1}{6} + 4\cdot\frac{1}{6} + 5\cdot\frac{1}{6} + 6\cdot\frac{1}{6}$$

$$= \frac{7}{2}$$

Ahora,

$$\mu_2' = E(X^2) = 1^2\cdot\frac{1}{6} + 2^2\cdot\frac{1}{6} + 3^2\cdot\frac{1}{6} + 4^2\cdot\frac{1}{6} + 5^2\cdot\frac{1}{6} + 6^2\cdot\frac{1}{6}$$

$$= \frac{91}{6}$$

y se sigue que

$$\sigma^2 = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12} \qquad \blacktriangle$$

EJEMPLO 4.11

Con respecto al ejemplo 4.2, encuentre la desviación estándar de la variable aleatoria X.

Solución

En el ejemplo 4.2 mostramos que $\mu = E(X) = 0.4413$. Ahora

$$\mu_2' = E(X^2) = \frac{4}{\pi}\int_0^1 \frac{x^2}{1+x^2}\,dx$$

$$= \frac{4}{\pi}\int_0^1 \left(1 - \frac{1}{1+x^2}\right) dx$$

$$= \frac{4}{\pi} - 1$$

$$= 0.2732$$

y se sigue que

$$\sigma^2 = 0.2732 - (0.4413)^2 = 0.0785$$

y $\sigma = \sqrt{0.0785} = 0.2802$. $\blacktriangle$

El siguiente es otro teorema que reviste importancia para el trabajo relacionado con desviaciones estándar o varianzas.

TEOREMA 4.7 Si X tiene la varianza σ^2, entonces

$$\text{var}(aX + b) = a^2\sigma^2$$

La demostración de este teorema se le dejará al lector, pero señalemos los siguientes corolarios: para $a = 1$, encontramos que añadir una constante a los valores de una variable aleatoria, resulta en un desplazamiento de todos los valores de X a la izquierda o la derecha, no afecta de manera alguna la amplitud de la distribución ; para $b = 0$, encontramos que si los valores de una variable aleatoria se multiplican por una constante, la varianza se multiplica por el cuadrado de esa constante, lo que resulta en un cambio correspondiente en la amplitud de la distribución.

4.4 TEOREMA DE CHEBYSHEV

Para demostrar cómo σ o σ^2 es indicativo de la amplitud o dispersión de la distribución de una variable aleatoria, demostraremos ahora el siguiente teorema, llamado **teorema de Chebyshev** en honor del matemático ruso del siglo XIX, P. L Chebyshev. Aquí sólo lo probaremos para el caso continuo, dejamos el caso discreto como un ejercicio.

TEOREMA 4.8 (*Teorema de Chebyshev*) Si μ y σ son la media y la desviación estándar de una variable aleatoria X, entonces para cualquier constante positiva k la probabilidad es *al menos* $1 - \frac{1}{k^2}$ que X asumirá un valor dentro de k desviaciones estándar de la media; simbólicamente

$$P(|X - \mu| < k\sigma) \geq 1 - \frac{1}{k^2}$$

Demostración. De acuerdo con las definiciones 4.4 y 4.5, escribimos

$$\sigma^2 = E[(X - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 \cdot f(x)\, dx$$

Entonces, al dividir la integral en tres partes como se muestra en la figura 4.2, obtenemos

$$\sigma^2 = \int_{-\infty}^{\mu - k\sigma} (x - \mu)^2 \cdot f(x)\, dx + \int_{\mu - k\sigma}^{\mu + k\sigma} (x - \mu)^2 \cdot f(x)\, dx$$
$$+ \int_{\mu + k\sigma}^{\infty} (x - \mu)^2 \cdot f(x)\, dx$$

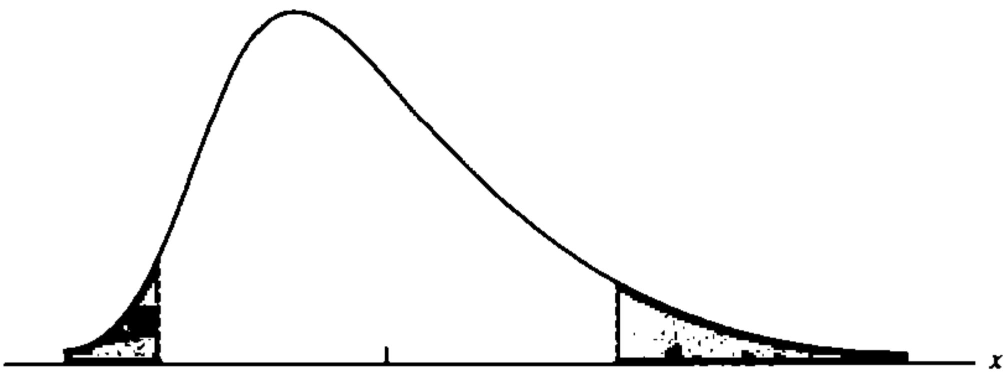

Figura 4.2 Diagrama para demostración del teorema de Chebyshev.

Puesto que el integrando $(x - \mu)^2 \cdot f(x)$ es no negativo, podemos formar la desigualdad

$$\sigma^2 \geqq \int_{-\infty}^{\mu - k\sigma} (x - \mu)^2 \cdot f(x)\, dx + \int_{\mu + k\sigma}^{\infty} (x - \mu)^2 \cdot f(x)\, dx$$

al borrar la segunda integral. Por consiguiente, puesto que $(x - \mu)^2 \geqq k^2\sigma^2$ para $x \leqq \mu - k\sigma$ o $x \geqq \mu + k\sigma$ se sigue que

$$\sigma^2 \geqq \int_{-\infty}^{\mu - k\sigma} k^2\sigma^2 \cdot f(x)\, dx + \int_{\mu + k\sigma}^{\infty} k^2\sigma^2 \cdot f(x)\, dx$$

y de ahí que

$$\frac{1}{k^2} \geqq \int_{-\infty}^{\mu - k\sigma} f(x)\, dx + \int_{\mu + k\sigma}^{\infty} f(x)\, dx$$

siempre y cuando $\sigma^2 \neq 0$. Puesto que la suma de las dos integrales en el lado derecho es la probabilidad de que X asumirá un valor menor que o igual a $\mu - k\sigma$ o mayor que o igual a $\mu + k\sigma$, hemos demostrado así que

$$P(|X - \mu| \geqq k\sigma) \leqq \frac{1}{k^2}$$

y se sigue que

$$P(|X - \mu| < k\sigma) \geqq 1 - \frac{1}{k^2} \qquad \blacktriangledown$$

Por ejemplo, la probabilidad es al menos $1 - \frac{1}{2^2} = \frac{3}{4}$ que una variable aleatoria X asumirá un valor dentro de dos desviaciones estándar de la media, la probabilidad es al menos $1 - \frac{1}{3^2} = \frac{8}{9}$ que asumirá un valor dentro de tres desviaciones estándar de la media y la probabilidad es al menos $1 - \frac{1}{5^2} = \frac{24}{25}$ que asumirá un valor dentro de cinco desviaciones estándar de la media. Es en este sentido que σ controla la amplitud o dispersión de la distribución de una variable aleatoria. Claramente, la probabilidad dada por el teorema de Chebyshev es solamente un límite inferior; si la probabilidad de que una variable aleatoria dada asuma un valor dentro de k desviaciones estándar de la media sea realmente mayor que $1 - \frac{1}{k^2}$ y, si es así, no podemos decir por cuánto, pero el teorema de Chebyshev nos asegura que esta probabilidad no puede ser menor que $1 - \frac{1}{k^2}$. Sólo cuando se conoce la distribución de una variable aleatoria podemos calcular la probabilidad exacta.

EJEMPLO 4.12

Si la densidad de probabilidad de X está dada por

$$f(x) = \begin{cases} 630x^4(1-x)^4 & \text{para } 0 < x < 1 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

encuentre la probabilidad de que asumirá un valor dentro de dos desviaciones estándar de la media y compare esta probabilidad con el límite inferior proporcionado por el teorema de Chebyshev.

Solución

La integración directa nos muestra que $\mu = \frac{1}{2}$ y $\sigma^2 = \frac{1}{44}$, de manera que $\sigma = \sqrt{1/44}$ o aproximadamente 0.15. Así, la probabilidad de que X asumirá un valor dentro de dos desviaciones estándar de la media es la probabilidad de que asumirá un valor entre 0.20 y 0.80, esto es,

$$\begin{aligned} P(0.20 < X < 0.80) &= \int_{0.20}^{0.80} 630x^4(1-x)^4\,dx \\ &= 0.96 \end{aligned}$$

Observe que la aseveración "la probabilidad es 0.96" es una aseveración más firme que "la probabilidad es al menos 0.75", la que es proporcionada por el teorema de Chebyshev. ▲

4.5 FUNCIONES GENERATRICES DE MOMENTOS

Aunque los momentos de la mayoría de las distribuciones se pueden determinar directamente al evaluar las integrales o sumas necesarias, un procedimiento alternativo algunas veces proporciona simplificaciones considerables. Esta técnica utiliza las **funciones generatrices de momentos**.

DEFINICIÓN 4.6 La función generatriz de momentos de una variable aleatoria X, donde existe, está dada por

$$M_X(t) = E(e^{tX}) = \sum_x e^{tx} \cdot f(x)$$

cuando X es discreta y

$$M_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} \cdot f(x)\, dx$$

cuando X es continua.

La variable independiente es t, y por lo general estamos interesados en los valores de t en la cercanía de 0.

Para explicar por qué nos referimos a esta función como una función "generatriz de momentos", sustituyamos por e^{tx} su expansión en la serie de Maclaurin, esto es,

$$e^{tx} = 1 + tx + \frac{t^2x^2}{2!} + \frac{t^3x^3}{3!} + \cdots + \frac{t^rx^r}{r!} + \cdots$$

Para el caso discreto, obtenemos así

$$\begin{aligned} M_X(t) &= \sum_x \left[1 + tx + \frac{t^2x^2}{2!} + \cdots + \frac{t^rx^r}{r!} + \cdots\right] f(x) \\ &= \sum_x f(x) + t \cdot \sum_x xf(x) + \frac{t^2}{2!} \cdot \sum_x x^2 f(x) + \cdots + \frac{t^r}{r!} \cdot \sum_x x^r f(x) + \cdots \\ &= 1 + \mu t + \mu_2' \cdot \frac{t^2}{2!} + \cdots + \mu_r' \cdot \frac{t^r}{r!} + \cdots \end{aligned}$$

y se puede ver que en la serie de Maclaurin de la función generatriz de momentos de X el coeficiente de $\frac{t^r}{r!}$ es μ_r', el $r$ésimo momento alrededor del origen. En el caso continuo, el argumento es el mismo.

EJEMPLO 4.13

Encuentre la función generatriz de momentos de la variable aleatoria cuya densidad de probabilidad está dada por

$$f(x) = \begin{cases} e^{-x} & \text{para } x > 0 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

y úsela para encontrar una expresión para μ_r'.

Solución

Por definición

$$
\begin{aligned}
M_X(t) = E(e^{tX}) &= \int_0^\infty e^{tx} \cdot e^{-x}\, dx \\
&= \int_0^\infty e^{-x(1-t)}\, dx \\
&= \frac{1}{1-t} \qquad \text{para } t < 1
\end{aligned}
$$

Como es bien sabido, cuando $|t| < 1$ la serie de Maclaurin para esta función generatriz de momentos es

$$
\begin{aligned}
M_X(t) &= 1 + t + t^2 + t^3 + \cdots + t^r + \cdots \\
&= 1 + 1! \cdot \frac{t}{1!} + 2! \cdot \frac{t^2}{2!} + 3! \cdot \frac{t^3}{3!} + \cdots + r! \cdot \frac{t^r}{r!} + \cdots
\end{aligned}
$$

y por tanto para $r = 0, 1, 2, \ldots$. ▲

La dificultad principal al usar la serie de Maclaurin de una función generatriz de momentos para determinar los momentos de una variable aleatoria generalmente *no* es encontrar la función generatriz de momentos, sino expandirla en la serie de Maclaurin. Si sólo estamos interesados en los pocos primeros momentos de una variable aleatoria, digamos, μ_1' y μ_2', generalmente podemos simplificar su determinación mediante el siguiente teorema.

TEOREMA 4.9

$$
\left.\frac{d^r M_X(t)}{dt^r}\right|_{t=0} = \mu_r'
$$

Esto se sigue del hecho que si una función se expande como una serie exponencial en t, el coeficiente de $\frac{t^r}{r!}$ es la résima derivada de la función con respecto a t en $t = 0$.

EJEMPLO 4.14

Dado que X tiene la distribución de probabilidad $f(x) = \frac{1}{8}\binom{3}{x}$ para $x = 0, 1, 2$ y 3, encuentre la función generatriz de momentos de esta variable aleatoria y úsela para determinar μ_1' y μ_2'.

Solución

De acuerdo a la definición 4.6

$$M_X(t) = E(e^{tX}) = \frac{1}{8}\cdot\sum_{x=0}^{3} e^{tx}\binom{3}{x}$$

$$= \frac{1}{8}(1 + 3e^t + 3e^{2t} + e^{3t})$$

$$= \frac{1}{8}(1 + e^t)^3$$

Entonces, por el teorema 4.9

$$\mu_1' = M_X'(0) = \frac{3}{8}(1 + e^t)^2 e^t\bigg|_{t=0} = \frac{3}{2}$$

y

$$\mu_2' = M_X''(0) = \frac{3}{4}(1 + e^t)e^{2t} + \frac{3}{8}(1 + e^t)^2 e^t\bigg|_{t=0} = 3 \qquad \blacktriangle$$

A menudo el trabajo que implica el uso de funciones generatrices de momentos se puede simplificar al usar el siguiente teorema.

> **TEOREMA 4.10** Si a y b son constantes, entonces
>
> 1. $M_{X+a}(t) = E[e^{(X+a)t}] = e^{at}\cdot M_X(t)$;
> 2. $M_{bX}(t) = E(e^{bXt}) = M_X(bt)$;
> 3. $M_{\frac{X+a}{b}}(t) = E[e^{\left(\frac{X+a}{b}\right)t}] = e^{\frac{a}{b}t}\cdot M_X\left(\frac{t}{b}\right)$.

La demostración de este teorema se deja al lector en el ejercicio 4.46. Como veremos más adelante, la primera parte del teorema es de especial importancia cuando $a = -\mu$, y la tercera parte es de especial importancia cuando $a = -\mu$ y $b = \sigma$, en cuyo caso

$$M_{\frac{X-\mu}{\sigma}}(t) = e^{-\frac{\mu t}{\sigma}}\cdot M_X\left(\frac{t}{\sigma}\right)$$

EJERCICIOS

4.25 Con respecto a la definición 4.4, demuestre que $\mu_0 = 1$ y que $\mu_1 = 0$ para cualquier variable aleatoria para la que exista $E(X)$.

4.26 Encuentre μ, μ_2' y σ^2 para la variable aleatoria X que tenga la distribución de probabilidad $f(x) = \frac{1}{2}$ para $x = -2$ y $x = 2$.

4.27 Encuentre μ, μ_2' y σ^2 para la variable aleatoria X que tiene la densidad de probabilidad

$$f(x) = \begin{cases} \dfrac{x}{2} & \text{para } 0 < x < 2 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

4.28 Encuentre μ_r' y σ^2 para la variable aleatoria X que tiene la densidad de probabilidad

$$f(x) = \begin{cases} \dfrac{1}{\ln 3} \cdot \dfrac{1}{x} & \text{para } 1 < x < 3 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

4.29 Demuestre el teorema 4.7.

4.30 Con respecto al ejercicio 4.8, encuentre la varianza de $g(X) = 2X + 3$.

4.31 Si la variable aleatoria X tiene la media μ y la desviación estándar σ, demuestre que la variable aleatoria Z cuyos valores están relacionados con los de X por medio de la ecuación $z = \dfrac{x - \mu}{\sigma}$ tiene $E(Z) = 0$ y $\text{var}(Z) = 1$

Una distribución que tiene la media 0 y la varianza 1 se dice que está en **forma normal**, y cuando efectuamos el cambio de variable anterior, decimos que estamos **normalizando** la distribución de X.

4.32 Si la densidad de probabilidad de X está dada por

$$f(x) = \begin{cases} 2x^{-3} & \text{para } x > 1 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

verifique si existen su media y su varianza.

4.33 Demuestre que

$$\mu_r = \mu_r' - \binom{r}{1}\mu_{r-1}' \cdot \mu + \cdots + (-1)^i\binom{r}{i}\mu_{r-i}' \cdot \mu^i + \cdots$$
$$+ (-1)^{r-1}(r - 1) \cdot \mu^r$$

para $r = 1, 2, 3, \ldots$, y use esta fórmula para expresar μ_3 y μ_4 en términos de momentos alrededor del origen.

4.34 La **simetría** o **asimetría** (falta de simetría) de una distribución a menudo se mide por la cantidad

$$\alpha_3 = \frac{\mu_3}{\sigma^3}$$

Use la fórmula para μ_3 obtenida en el ejercicio 4.33 para determinar α_3 para cada una de las siguientes distribuciones (las cuales tienen medias y desviaciones estándar iguales):

(a) $f(1) = 0.05$, $f(2) = 0.15$, $f(3) = 0.30$, $f(4) = 0.30$, $f(5) = 0.15$ y $f(6) = 0.05$;

(b) $f(1) = 0.05$, $f(2) = 0.20$, $f(3) = 0.15$, $f(4) = 0.45$, $f(5) = 0.10$ y $f(6) = 0.05$.

También dibuje los histogramas de las dos distribuciones y advierta que mientras la primera es simétrica, la segunda tiene una "cola" en el lado izquierdo y se dice que es **negativamente asimétrica**.

4.35 El grado en el cual una distribución sea puntiaguda o aplanada, también se conoce como la **curtosis** de la distribución, a menudo se mide por medio de la cantidad

$$\alpha_4 = \frac{\mu_4}{\sigma^4}$$

Use la fórmula para μ_4 obtenida en el ejercicio 4.33 para encontrar α_4 para cada una de las siguientes distribuciones simétricas, de la cual la primera es más puntiaguda (pico más angosto) que la segunda:

(a) $f(-3) = 0.06$, $f(-2) = 0.09$, $f(-1) = 0.10$, $f(0) = 0.50$, $f(1) = 0.10$, $f(2) = 0.09$ y $f(3) = 0.06$;

(b) $f(-3) = 0.04$, $f(-2) = 0.11$, $f(-1) = 0.20$, $f(0) = 0.30$, $f(1) = 0.20$, $f(2) = 0.11$ y $f(3) = 0.04$.

4.36 Repita los pasos usados en la demostración del teorema 4.8 para probar el teorema de Chebyshev para una variable aleatoria discreta X.

4.37 Demuestre que si X es una variable aleatoria con la media μ para la cual $f(x) = 0$ para $x < 0$, entonces para cualquier constante positiva a,

$$P(X \geqq a) \leqq \frac{\mu}{a}$$

Esta desigualdad se llama **desigualdad de Markov**, y la damos aquí principalmente porque nos lleva a una demostración alternativa, relativamente simple del teorema de Chebyshev.

4.38 Use la desigualdad del ejercicio 4.37 para probar el teorema de Chebyshev [*Sugerencia*: sustituya $(X - \mu)^2$ en vez de X.]

4.39 ¿Cuál es el valor más pequeño de k en el teorema de Chebyshev para el que la probabilidad de que una variable aleatoria asuma un valor entre $\mu - k\sigma$ y $\mu + k\sigma$ sea

(a) al menos 0.95;

(b) al menos 0.99?

4.40 Si hacemos $k\sigma = c$ en el teorema de Chebyshev, ¿qué afirma este teorema sobre la probabilidad de que una variable aleatoria asuma un valor entre $\mu - c$ y $\mu + c$?

4.41 Encuentre la función generatriz de momentos de la variable aleatoria discreta X que tiene la distribución de probabilidad

$$f(x) = 2\left(\frac{1}{3}\right)^x \qquad \text{para } x = 1, 2, 3, \ldots$$

y úsela para determinar los valores de μ_1' y μ_2'.

4.42 Encuentre la función generatriz de momentos de la variable aleatoria continua X cuya densidad de probabilidad está dada por

$$f(x) = \begin{cases} 1 & \text{para } 0 < x < 1 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

y úsela para encontrar μ_1', μ_2' y σ^2.

4.43 Si hacemos $R_X(t) = \ln M_X(t)$, demuestre que $R_X'(0) = \mu$ y $R_X''(0) = \sigma^2$. También, use estos resultados para encontrar la media y la varianza de una variable aleatoria X que tenga la función generatriz de momentos

$$M_X(t) = e^{4(e^t-1)}$$

4.44 Explique por qué no puede haber una variable aleatoria para la cual $M_X(t) = \dfrac{t}{1-t}$.

4.45 Demuestre que si una variable aleatoria tiene la densidad de probabilidad

$$f(x) = \frac{1}{2}e^{-|x|} \qquad \text{para } -\infty < x < \infty$$

su función generatriz de momentos está dada por

$$M_X(t) = \frac{1}{1-t^2}$$

4.46 Con respecto al ejercicio 4.45, encuentre la varianza de una variable aleatoria

(a) al expandir la función generatriz de momentos como una serie infinita y leer de ella los coeficientes necesarios;

(b) al usar el teorema 4.9.

4.47 Demuestre las tres partes del teorema 4.10.

4.48 Dada la función generatriz de momentos $M_X(t) = e^{3t+8t^2}$, encuentre la función generatriz de momentos de la variable aleatoria $Z = \frac{1}{4}(X-3)$, y úsela para determinar la media y la varianza de Z.

APLICACIONES

4.49 Con respecto al ejemplo 4.1, encuentre la varianza del número de aparatos de televisión con cables blancos.

4.50 El tiempo que toma servir a una persona en un restaurante dado es una variable aleatoria con la densidad de probabilidad

$$f(x) = \begin{cases} \frac{1}{4}e^{-\frac{x}{4}} & \text{para } x > 0 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

Encuentre la media y la varianza de esta variable aleatoria

4.51 Con respecto al ejercicio 3.51, encuentre la media y la varianza de la variable aleatoria en cuestión.

4.52 Con respecto al ejercicio 3.20, encuentre la media y la varianza de la variable aleatoria V.

4.53 Lo que sigue son algunas aplicaciones de la desigualdad de Markov del ejercicio 4.37:

(a) La puntuación que un estudiante de tercer año de secundaria obtiene en la parte verbal de la prueba PSAT/NMSQT se puede considerar como los valores de una variable aleatoria con la media $\mu = 41$. Encuentre el límite superior de la probabilidad de que uno de los estudiantes obtenga una puntuación de 65 o más.

(b) El peso de ciertos animales se puede considerar como una variable aleatoria con una media de 212 gramos. Si ninguno de los animales pesa menos de 165 gramos, encuentre un límite superior a la probabilidad de que un animal así pese al menos 250 gramos.

4.54 El número de licencias de matrimonio expedidas en cierta ciudad durante el mes de junio se puede considerar como una variable aleatoria con $\mu = 124$ y $\sigma = 7.5$. De acuerdo con el teorema de Chebyshev, ¿con qué probabilidad podemos afirmar que ahí se emitirán entre 64 y 184 licencias de matrimonio durante el mes de junio?

4.55 Un estudio del valor alimenticio de cierta clase de pan muestra que la cantidad de tiamina (vitamina B_1) en una rebanada se puede considerar como una variable aleatoria con $\mu = 0.260$ miligramos y $\sigma = 0.005$ miligramos. De acuerdo al teorema de Chebyshev, ¿entre qué valores debe estar el contenido de tiamina de

(a) al menos $\frac{35}{36}$ de todas las rebanadas de este pan;

(b) al menos $\frac{143}{144}$ de todas las rebanadas de este pan?

4.56 Con respecto al ejercicio 4.50, ¿qué podemos afirmar sobre la cantidad de tiempo que toma servir a una persona en el restaurante dado si usamos el teorema de Chebyshev con $k = 1.5$? ¿Cuál es la probabilidad correspondiente con cuatro decimales?

4.6 MOMENTOS PRODUCTO

Para continuar el examen de la sección 4.3, presentaremos ahora los **momentos producto** de dos variables aleatorias.

DEFINICIÓN 4.7 El $r$ésimo y el $s$ésimo **momentos producto alrededor del origen** de las variables aleatorias X y Y, denotadas por $\mu'_{r,s}$, es el valor esperado de X^rY^s; simbólicamente,

$$\mu'_{r,s} = E(X^rY^s) = \sum_x \sum_y x^r y^s \cdot f(x, y)$$

para $r = 0, 1, 2, \ldots$ y $s = 0, 1, 2, \ldots$ cuando X y Y son discretas, y

$$\mu'_{r,s} = E(X^rY^s) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^ry^s\cdot f(x, y)\,dx\,dy$$

donde X y Y son continuas.

En el caso discreto, la doble suma se extiende en el intervalo conjunto completo de las dos variables aleatorias. Advierta que $\mu'_{1,0} = E(X)$, el cual denotamos aquí con μ_X, y que $\mu'_{0,1} = E(Y)$, el cual denotamos aquí con μ_Y.

Análoga a la definición 4.4, definamos ahora los momentos producto alrededor de sus medias respectivas.

DEFINICIÓN 4.8 El $r$ésimo y $s$ésimo **momentos producto alrededor de la media** de las variables aleatorias X y Y, denotadas por $\mu_{r,s}$, es el valor esperado de $(X - \mu_X)^r(Y - \mu_Y)^s$; simbólicamente,

$$\mu_{r,s} = E[(X - \mu_X)^r(Y - \mu_Y)^s]$$
$$= \sum_x \sum_y (x - \mu_X)^r(y - \mu_Y)^s\cdot f(x, y)$$

para $r = 0, 1, 2, \ldots$ y $s = 0, 1, 2, \ldots$ cuando X y Y son discretas, y

$$\mu_{r,s} = E[(X - \mu_X)^r(Y - \mu_Y)^s]$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x - \mu_X)^r(y - \mu_Y)^s\cdot f(x, y)\,dx\,dy$$

cuando X y Y son continuas.

En estadística, $\mu_{1,1}$ es de especial importancia porque es indicativa de la relación, si es que la hay, entre los valores X y Y; así, se le da un símbolo especial y un nombre especial.

DEFINICIÓN 4.9 $\mu_{1,1}$ se llama la **covarianza** de X y Y, y se denota con σ_{XY}, cov(X, Y) o $C(X, Y)$.

Observe que si existe una probabilidad alta de que valores grandes de X vayan con valores grandes de Y y valores pequeños de X con valores pequeños de Y, la covarianza será positiva; si hay una alta probabilidad de que valores grandes de X vayan con valores pequeños de Y, y viceversa, la covarianza será negativa. Es en este sentido que la covarianza mide la relación, o asociación, entre los valores de X y Y.

Probemos ahora el siguiente resultado, análogo al teorema 4.6, que es útil en determinar realmente las covarianzas.

TEOREMA 4.11

$$\sigma_{XY} = \mu'_{1,1} - \mu_X\mu_Y$$

Demostración. Al usar los diversos teoremas sobre los valores esperados, podemos escribir

$$\begin{aligned}\sigma_{XY} &= E[(X-\mu_X)(Y-\mu_Y)]\\ &= E(XY - X\mu_Y - Y\mu_X + \mu_X\mu_Y)\\ &= E(XY) - \mu_Y E(X) - \mu_X E(Y) + \mu_X\mu_Y\\ &= E(XY) - \mu_Y\mu_X - \mu_X\mu_Y + \mu_X\mu_Y\\ &= \mu'_{1,1} - \mu_X\mu_Y \qquad \blacktriangledown\end{aligned}$$

EJEMPLO 4.15

En el ejemplo 3.20, las probabilidades marginal y conjunta de X y Y, los números de cápsulas de aspirinas y sedantes entre dos cápsulas sacadas al azar de un frasco que contiene tres aspirinas, dos sedantes y cuatro cápsulas laxantes, se registraron como sigue:

		x			
		0	1	2	
y	0	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{12}$	$\frac{7}{12}$
	1	$\frac{2}{9}$	$\frac{1}{6}$		$\frac{7}{18}$
	2	$\frac{1}{36}$			$\frac{1}{36}$
		$\frac{5}{12}$	$\frac{1}{2}$	$\frac{1}{12}$	

Encuentre la covarianza de X y Y.

Solución

Al referirnos a las probabilidades conjuntas dadas aquí, obtenemos

$$\begin{aligned}\mu'_{1,1} &= E(XY)\\ &= 0\cdot 0\cdot\frac{1}{6} + 0\cdot 1\cdot\frac{2}{9} + 0\cdot 2\cdot\frac{1}{36} + 1\cdot 0\cdot\frac{1}{3} + 1\cdot 1\cdot\frac{1}{6} + 2\cdot 0\cdot\frac{1}{12}\\ &= \frac{1}{6}\end{aligned}$$

y al usar las probabilidades marginales, obtenemos

$$\mu_X = E(X) = 0 \cdot \frac{5}{12} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{12} = \frac{2}{3}$$

y

$$\mu_Y = E(Y) = 0 \cdot \frac{7}{12} + 1 \cdot \frac{7}{18} + 2 \cdot \frac{1}{36} = \frac{4}{9}$$

Se sigue que

$$\sigma_{XY} = \frac{1}{6} - \frac{2}{3} \cdot \frac{4}{9} = -\frac{7}{54}$$

El resultado negativo sugiere que mientras saquemos más tabletas de aspirinas, sacaremos menos tabletas de sedante y viceversa, y esto por supuesto es razonable. ▲

EJEMPLO 4.16

Encuentre la covarianza de las variables aleatorias cuya densidad de probabilidad está dada por

$$f(x, y) = \begin{cases} 2 & \text{para } x > 0, y > 0, x + y < 1 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

Solución

Al evaluar las integrales necesarias, obtenemos

$$\mu_X = \int_0^1 \int_0^{1-x} 2x \, dy \, dx = \frac{1}{3}$$

$$\mu_Y = \int_0^1 \int_0^{1-x} 2y \, dy \, dx = \frac{1}{3}$$

y

$$\sigma'_{1,1} = \int_0^1 \int_0^{1-x} 2xy \, dy \, dx = \frac{1}{12}$$

Se sigue que

$$\sigma_{XY} = \frac{1}{12} - \frac{1}{3} \cdot \frac{1}{3} = -\frac{1}{36}$$ ▲

Por lo que toca a la relación entre X y Y, observe que si X y Y son independientes, su covarianza es cero; simbólicamente,

TEOREMA 4.12 Si X y Y son independientes, entonces $E(XY) = E(X) \cdot E(Y)$ y $\sigma_{XY} = 0$.

Demostración. Para el caso discreto tenemos, por definición,

$$E(XY) = \sum_x \sum_y xy \cdot f(x, y)$$

Puesto que X y Y son independientes, podemos escribir $f(x, y) = g(x) \cdot h(y)$, donde $g(x)$ y $h(y)$ donde $g(x)$ y $h(y)$ son los valores de las distribuciones marginales de X y Y, y obtenemos

$$\begin{aligned} E(XY) &= \sum_x \sum_y xy \cdot g(x)h(y) \\ &= \left[\sum_x x \cdot g(x)\right]\left[\sum_y y \cdot h(y)\right] \\ &= E(X) \cdot E(Y) \end{aligned}$$

Por tanto,

$$\begin{aligned} \sigma_{XY} &= \mu'_{1,1} - \mu_X \mu_Y \\ &= E(X) \cdot E(Y) - E(X) \cdot E(Y) \\ &= 0 \qquad \blacktriangledown \end{aligned}$$

Es interesante señalar que la independencia de dos variables aleatorias implica una covarianza cero, pero una covarianza cero no implica necesariamente su independencia. Esto se ilustra con el siguiente ejemplo (véanse también los ejercicios 4.61 y 4.62).

EJEMPLO 4.17

Si la distribución de probabilidad conjunta de X y Y está dada por

		$x = -1$	$x = 0$	$x = 1$	
y	-1	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{2}{3}$
	0	0	0	0	0
	1	$\frac{1}{6}$	0	$\frac{1}{6}$	$\frac{1}{3}$
		$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	

demuestre que su covarianza es cero aun cuando las dos variables aleatorias no son independientes.

Solución

Al usar las probabilidades mostradas en los márgenes, obtenemos

$$\mu_X = (-1)\cdot\frac{1}{3} + 0\cdot\frac{1}{3} + 1\cdot\frac{1}{3} = 0$$

$$\mu_Y = (-1)\cdot\frac{2}{3} + 0\cdot 0 + 1\cdot\frac{1}{3} = -\frac{1}{3}$$

y

$$\mu'_{1,1} = (-1)(-1)\cdot\frac{1}{6} + 0(-1)\cdot\frac{1}{3} + 1(-1)\cdot\frac{1}{6} + (-1)1\cdot\frac{1}{6} + 1\cdot 1\cdot\frac{1}{6}$$
$$= 0$$

Así, $\sigma_{XY} = 0 - 0(-\frac{1}{3}) = 0$, la covarianza es cero, pero las dos variables aleatorias no son independientes. Por ejemplo, $f(x, y) \neq g(x)\cdot h(y)$ para $x = -1$ y $y = -1$. ▲

Los momentos producto también se pueden definir para el caso donde hay más de dos variables aleatorias. Aquí simplemente enunciemos el resultado importante que

> **TEOREMA 4.13** Si $X_1, X_2, \ldots, X_n$ son independientes, entonces
> $$E(X_1 X_2 \cdot \ldots \cdot X_n) = E(X_1)\cdot E(X_2)\cdot \ldots \cdot E(X_n)$$

Ésta es una generalización de la primera parte del teorema 4.12; de hecho, la demostración de este teorema, basado en la definición 3.14, es esencialmente como la de la primera parte del teorema 4.12.

4.7 MOMENTOS DE COMBINACIONES LINEALES DE VARIABLES ALEATORIAS

En esta sección derivaremos expresiones para la media y la varianza de una combinación lineal de n variables aleatorias y la covarianza de dos combinaciones lineales de n variables aleatorias. Las aplicaciones de estos resultados se tratarán más adelante en nuestro examen de la teoría del muestreo y los problemas de inferencia estadística.

> **TEOREMA 4.14** Si $X_1, X_2, \ldots, X_n$ son variables aleatorias y
> $$Y = \sum_{i=1}^{n} a_i X_i$$
> donde $a_1, a_2, \ldots, a_n$ son constantes, entonces

$$E(Y) = \sum_{i=1}^{n} a_i E(X_i)$$

y

$$\text{var}(Y) = \sum_{i=1}^{n} a_i^2 \cdot \text{var}(X_i) + 2 \sum_{i<j}\sum a_i a_j \cdot \text{cov}(X_i X_j)$$

donde la doble suma se extiende para todos los valores de i y j, desde 1 hasta n, para los que $i < j$.

Demostración. Partiendo del teorema 4.5 con $g_i(X_1, X_2, \ldots, X_k) = X_i$ para $i = 0, 1, 2, \ldots, n$, se sigue inmediatamente que

$$E(Y) = E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i E(X_i)$$

y esto demuestra la primera parte del teorema. Para obtener la expresión de la varianza de Y, escribamos μ_i para $E(X_i)$ de manera que obtenemos

$$\text{var}(Y) = E([Y - E(Y)]^2) = E\left\{\left[\sum_{i=1}^{n} a_i X_i - \sum_{i=1}^{n} a_i E(X_i)\right]^2\right\}$$

$$= E\left\{\left[\sum_{i=1}^{n} a_i (X_i - \mu_i)\right]^2\right\}$$

Entonces, al expandir por medio del teorema multinomial, de acuerdo al cual $(a + b + c + d)^2$, por ejemplo, es igual a $a^2 + b^2 + c^2 + d^2 + 2ab + 2ac + 2ad + 2bc + 2bd + 2cd$, y refiriéndonos otra vez al teorema 4.5, obtenemos

$$\text{var}(Y) = \sum_{i=1}^{n} a_i^2 E[(X_i - \mu_i)^2] + 2\sum_{i<j}\sum a_i a_j E[(X_i - \mu_i)(X_j - \mu_j)]$$

$$= \sum_{i=1}^{n} a_i^2 \cdot \text{var}(X_i) + 2\sum_{i<j}\sum a_i a_j \cdot \text{cov}(X_i, X_j)$$

Advierta que tácitamente nos valimos del hecho que $\text{cov}(X_i, X_j) = \text{cov}(X_j, X_i)$. ▼

Puesto que $\text{cov}(X_i, X_j) = 0$ cuando X_i y X_j son independientes, se sigue inmediatamente que

COROLARIO Si las variables aleatorias $X_1, X_2, \ldots, X_n$ son independientes y $Y = \sum_{i=1}^{n} a_i X_i$, entonces

$$\text{var}(Y) = \sum_{i=1}^{n} a_i^2 \cdot \text{var}(X_i)$$

EJEMPLO 4.18

Si las variables aleatorias X, Y y Z tienen las medias $\mu_X = 2$, $\mu_Y = -3$ y $\mu_Z = 4$, las varianzas $\sigma_X^2 = 1$, $\sigma_Y^2 = 5$ y $\sigma_Z^2 = 2$, y las covarianzas $\text{cov}(X, Y) = -2$, $\text{cov}(X, Z) = -1$ y $\text{cov}(Y, Z) = 1$, encuentre la media y la varianza de $W = 3X - Y + 2Z$.

Solución

Por medio del teorema 4.14, obtenemos

$$\begin{aligned} E(W) &= E(3X - Y + 2Z) \\ &= 3E(X) - E(Y) + 2E(Z) \\ &= 3 \cdot 2 - (-3) + 2 \cdot 4 \\ &= 17 \end{aligned}$$

y

$$\begin{aligned} \text{var}(W) &= 9\ \text{var}(X) + \text{var}(Y) + 4\ \text{var}(Z) - 6\ \text{cov}(X, Y) \\ &\quad + 12\ \text{cov}(X, Z) - 4\ \text{cov}(Y, Z) \\ &= 9 \cdot 1 + 5 + 4 \cdot 2 - 6(-2) + 12(-1) - 4 \cdot 1 \\ &= 18 \qquad \blacktriangle \end{aligned}$$

Lo que sigue es otro teorema importante acerca de combinaciones lineales de variables aleatorias; se refiere a la covarianza de dos combinaciones lineales de n variables aleatorias.

TEOREMA 4.15 Si $X_1, X_2, \ldots, X_n$ son variables aleatorias y

$$Y_1 = \sum_{i=1}^{n} a_i X_i \quad \text{y} \quad Y_2 = \sum_{i=1}^{n} b_i X_i$$

donde $a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_n$ son constantes, entonces

$$\text{cov}(Y_1, Y_2) = \sum_{i=1}^{n} a_i b_i \cdot \text{var}(X_i) + \sum_{i<j}\sum (a_i b_j + a_j b_i) \cdot \text{cov}(X_i, X_j)$$

La demostración de este teorema, que es muy similar a la del teorema 4.14, se dejará al lector en el ejercicio 4.67.

Puesto que $\text{cov}(X_i, X_j) = 0$ cuando X_i y X_j son independientes, se sigue inmediatamente que

> **COROLARIO** Si las variables aleatorias $X_1, X_2, \ldots, X_n$ son independientes $Y_1 = \sum_{i=1}^{n} a_i X_i$ y $Y_2 = \sum_{i=1}^{n} b_i X_i$, entonces
>
> $$\text{cov}(Y_1, Y_2) = \sum_{i=1}^{n} a_i b_i \cdot \text{var}(X_i)$$

EJEMPLO 4.19

Si las variables aleatorias X, Y y Z tienen las medias $\mu_X = 3$, $\mu_Y = 5$ y $\mu_Z = 2$, las varianzas $\sigma_X^2 = 8$, $\sigma_Y^2 = 12$ y $\sigma_Z^2 = 18$, y $\text{cov}(X, Y) = 1$, $\text{cov}(X, Z) = -3$ y $\text{cov}(Y, Z) = 2$, encuentre la covarianza de

$$U = X + 4Y + 2Z \quad \text{y} \quad V = 3X - Y - Z$$

Solución

Por el teorema 4.15, obtenemos

$$\begin{aligned} \text{cov}(U, V) &= \text{cov}(X + 4Y + 2Z, 3X - Y - Z) \\ &= 3\,\text{var}(X) - 4\,\text{var}(Y) - 2\,\text{var}(Z) + 11\,\text{cov}(X, Y) \\ &\quad + 5\,\text{cov}(X, Z) - 6\,\text{cov}(Y, Z) \\ &= 3 \cdot 8 - 4 \cdot 12 - 2 \cdot 18 + 11 \cdot 1 + 5(-3) - 6 \cdot 2 \\ &= -76 \qquad \blacktriangle \end{aligned}$$

4.8 ESPERANZA CONDICIONAL

En la sección 3.7 obtuvimos las probabilidades condicionales al sumar los valores de las distribuciones de probabilidades condicionales o al integrar los valores de las densidades de probabilidades condicionales. **Las esperanzas condicionales** de variables aleatorias se definen de la misma manera en términos de sus distribuciones condicionales.

> **DEFINICIÓN 4.10** Si X es una variable aleatoria discreta y $f(x|y)$ es el valor de la distribución de probabilidad condicional de X dado $Y = y$ en x, la **esperanza condicional** de $u(X)$ dado $Y = y$ es
>
> $$E[u(X)|y] = \sum_{x} u(x) \cdot f(x|y)$$

De manera correspondiente, si X es una variable aleatoria continua y $f(x|y)$ es el valor de la densidad de probabilidad condicional de X dado $Y = y$ en x, la **esperanza condicional** de $u(X)$ dado $Y = y$ es

$$E[u(X)|y] = \int_{-\infty}^{\infty} u(x) \cdot f(x|y)\, dx$$

Expresiones similares basadas en la distribución o densidad de probabilidad condicional de Y dado $X = x$ definen la esperanza condicional de $v(Y)$ dado $X = x$.

Si hacemos $u(X) = X$ en la definición 4.10, obtenemos la **media condicional** de la variable aleatoria X dado $Y = y$, la cual denotamos por

$$\mu_{X|y} = E(X|y)$$

De manera correspondiente, la **varianza condicional** de X dado $Y = y$ es

$$\begin{aligned}\sigma^2_{X|y} &= E[(X - \mu_{X|y})^2|y] \\ &= E(X^2|y) - \mu^2_{X|y}\end{aligned}$$

donde $E(X^2|y)$ está dada por la definición 4.10 con $u(X) = X^2$. Al lector no le debe ser difícil generalizar la definición 4.10 para esperanzas condicionales que implican más de dos variables.

EJEMPLO 4.20

Con respecto al ejemplo 3.12, encuentre la media condicional de X dado $Y = 1$.

Solución

Al usar los resultados obtenidos en el ejemplo 3.23, esto es, $f(0|1) = \frac{4}{7}$, $f(1|1) = \frac{3}{7}$ y $f(2|1) = 0$, obtenemos

$$E(X|1) = 0 \cdot \frac{4}{7} + 1 \cdot \frac{3}{7} + 2 \cdot 0 = \frac{3}{7} \qquad \blacktriangle$$

EJEMPLO 4.21

Si la densidad de probabilidad conjunta de X y Y está dada por

$$f(x, y) = \begin{cases} \frac{2}{3}(x + 2y) & \text{para } 0 < x < 1, 0 < y < 1 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

encuentre la media condicional y la varianza condicional de X dado $Y = \frac{1}{2}$.

Solución

En el ejemplo 3.24 mostramos que para esas variables aleatorias la densidad condicional de X dado $Y = y$ es

$$f(x|y) = \begin{cases} \dfrac{2x + 4y}{1 + 4y} & \text{para } 0 < x < 1 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

de manera que

$$f\left(x\middle|\frac{1}{2}\right) = \begin{cases} \dfrac{2}{3}(x + 1) & \text{para } 0 < x < 1 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

Así, $\mu_{X|\frac{1}{2}}$ está dada por

$$E\left(X\middle|\frac{1}{2}\right) = \int_0^1 \frac{2}{3}x(x+1)\,dx$$

$$= \frac{5}{9}$$

A continuación encontramos

$$E\left(X^2\middle|\frac{1}{2}\right) = \int_0^1 \frac{2}{3}x^2(x+1)\,dx$$

$$= \frac{7}{18}$$

y se sigue que

$$\sigma^2_{X|\frac{1}{2}} = \frac{7}{18} - \left(\frac{5}{9}\right)^2 = \frac{13}{162} \qquad \blacktriangle$$

EJERCICIOS

4.57 Si X y Y tienen la distribución de probabilidad conjunta $f(x, y) = \frac{1}{4}$ para $x = -3$ y $y = -5$, $x = -1$ y $y = -1$, $x = 1$ y $y = 1$, y $x = 3$ y $y = 5$, encuentre $\text{cov}(X, Y)$.

4.58 Con respecto al ejercicio 3.56, encuentre la covarianza de X y Y.

4.59 Con respecto al ejercicio 3.22, encuentre la covarianza de X_1 y X_3.

4.60 Con respecto al ejercicio 3.94, encuentre $\text{cov}(X, Y)$.

4.61 Si X y Y tienen la distribución de probabilidad conjunta $f(-1, 0) = 0$, $f(-1, 1) = \frac{1}{4}$, $f(0, 0) = \frac{1}{6}$, $f(0, 1) = 0$, $f(1, 0) = \frac{1}{12}$ y $f(1, 1) = \frac{1}{2}$, demuestre que

(a) $\text{cov}(X, Y) = 0$;

(b) las dos variables aleatorias no son independientes.

4.62 Si la densidad de probabilidad de X está dada por

$$f(x) = \begin{cases} 1 + x & \text{para } -1 < x \leqq 0 \\ 1 - x & \text{para } 0 < x < 1 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

y $U = X$ y $V = X^2$, demostrar que

(a) $\text{cov}(U, V) = 0$;
(b) U y V son dependientes.

4.63 Para k variables aleatorias $X_1, X_2, \ldots, X_k$, los valores de la **función generatriz de momentos conjunta** están dados por

$$E\left(e^{t_1 X_1 + t_2 X_2 + \cdots + t_k X_k}\right)$$

(a) Demuestre para el caso discreto o para el caso continuo que la derivada parcial de la función generatriz de momentos conjunta con respecto a t_i en t_i a $t_1 = t_2 = \cdots = t_k = 0$ es $E(X_i)$.
(b) Demuestre para el caso discreto o para el caso continuo que la segunda derivada parcial de la función generatriz de momentos conjunta con respecto a t_i y t_j, $i \neq j$, en $t_1 = t_2 = \cdots = t_k = 0$ es $E(X_i X_j)$.
(c) Si dos variables aleatorias tienen la densidad conjunta dada por

$$f(x, y) = \begin{cases} e^{-x-y} & \text{para } x > 0, y > 0 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

encuentre su función generatriz de momentos conjunta y úsela para determinar los valores de $E(XY)$, $E(X)$, $E(Y)$ y $\text{cov}(X, Y)$.

4.64 Si X_1, X_2 y X_3 son independientes y tienen las medias 4, 9 y 3, y las varianzas 3, 7 y 5, encuentre la media y la varianza de
(a) $Y = 2X_1 - 3X_2 + 4X_3$;
(b) $Z = X_1 + 2X_2 - X_3$.

4.65 Repita ambas partes del ejercicio 4.64, sin hacer la suposición de independencia y en vez de ello use la información de que $\text{cov}(X_1, X_2) = 1$, $\text{cov}(X_2, X_3) = -2$ y $\text{cov}(X_1, X_3) = -3$.

4.66 Si la densidad de probabilidad conjunta de X y Y está dada por

$$f(x, y) = \begin{cases} \frac{1}{3}(x + y) & \text{para } 0 < x < 1, 0 < y < 2 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

encuentre la varianza de $W = 3X + 4Y - 5$.

4.67 Demuestre el teorema 4.15

4.68 Exprese $\text{var}(X + Y)$, $\text{var}(X - Y)$ y $\text{cov}(X + Y, X - Y)$ en términos de las varianzas y la covarianza de X y Y.

4.69 Si $\text{var}(X_1) = 5$, $\text{var}(X_2) = 4$, $\text{var}(X_3) = 7$, $\text{cov}(X_1, X_2) = 3$, $\text{cov}(X_1, X_3) = -2$, y X_2 y X_3 son independientes, encuentre la covarianza de $Y_1 = X_1 - 2X_2 + 3X_3$ y $Y_2 = -2X_1 + 3X_2 + 4X_3$.

4.70 Con respecto al ejercicio 4.64, encuentre $\text{cov}(Y, Z)$.

4.71 Con respecto al ejercicio 3.89, encuentre la media condicional y la varianza condicional de X dado $Y = -1$.

4.72 Con respecto al ejercicio 3.91, encuentre la esperanza condicional de la variable aleatoria $U = Z^2$ dado $X = 1$ y $Y = 2$.

4.73 Con respecto al ejercicio 3.94, encuentre la media condicional y la varianza condicional de Y dado $X = \frac{1}{4}$.

4.74 Con respecto al ejemplo 3.22 y el inciso (b) del ejercicio 3.98, encuentre el valor esperado de $X_2^2 X_3$ dado $X_1 = \frac{1}{2}$.

4.75 (a) Demuestre que la función de distribución condicional de la variable aleatoria continua X, dada por $a < X \leqq b$, está dada por

$$F(x|a < X \leqq b) = \begin{cases} 0 & \text{para } x \leqq a \\ \dfrac{F(x) - F(a)}{F(b) - F(a)} & \text{para } a < x \leqq b \\ 1 & \text{para } x > b \end{cases}$$

(b) Diferencie el resultado del inciso (a) con respecto a x para encontrar la densidad de probabilidad condicional de X dado $a < X \leqq b$, y demuestre que

$$E[u(X)|a < X \leqq b] = \frac{\int_a^b u(x)f(x)\,dx}{\int_a^b f(x)\,dx}$$

APLICACIONES

4.76 Una moneda de 25 centavos está torcida de manera que las probabilidades de caras y cruces son 0.40 y 0.60. Si se lanza dos veces, ¿cuál es la covarianza de Z, el número de caras obtenidas en el primer tiro, y W, el número total de caras obtenidas en los dos tiros de la moneda?

4.77 El diámetro interior de un tubo cilíndrico es una variable aleatoria con una media de 3 pulgadas y una desviación estándar de 0.02 pulgadas, el espesor del tubo es una variable aleatoria con una media de 0.3 pulgadas y una desviación estándar de 0.005 pulgadas, y las dos variables aleatorias son independientes. Encuentre la media y la desviación estándar del diámetro exterior del tubo.

4.78 La longitud de ciertos ladrillos es una variable aleatoria con una media de 8 pulgadas y una desviación estándar de 0.1 pulgadas, y el espesor del mortero entre dos ladrillos es una variable aleatoria con una media de 0.5 pulgadas y una desviación estándar de 0.03 pulgadas. ¿Cuál es la media y la desviación estándar de la longitud de una pared construida con 50 de estos ladrillos puestos lado a lado, si podemos suponer que todas las variables aleatorias implicadas son independientes?

4.79 Si obtener cara es un éxito cuando lanzamos una moneda, sacar un seis es un éxito cuando tiramos un dado, y sacar un as es un éxito cuando sacamos una carta de una baraja común de 52 cartas de juego, encuentre la media y la desviación estándar del número total de éxitos cuando

(a) lanzamos una moneda equilibrada, tiramos un dado balanceado y entonces sacamos una carta de una baraja bien barajada;

(b) lanzamos tres veces una moneda equilibrada, tiramos dos veces un dado balanceado, y entonces sacamos una carta de una baraja bien barajada.

4.80 Si lanzamos alternativamente una moneda equilibrada y una que está cargada de manera que la probabilidad de obtener cara sea 0.45, ¿cuál es la media y la desviación estándar del número de caras que obtenemos en 10 lanzamientos de estas monedas?

4.81 Con respecto al ejercicio 3.84 y el inciso (b) del ejercicio 3.103, encuentre el número esperado de textos de matemáticas dado en que no se selecciona ninguno de los textos de estadística.

4.82 Con respecto al ejercicio 3.107, ¿cuánto puede esperar se le reembolse un vendedor que ha gastado $12 en gasolina?

4.83 La cantidad de tiempo (en minutos) que el ejecutivo de cierta empresa habla por teléfono es una variable aleatoria que tiene la densidad de probabilidad

$$f(x) = \begin{cases} \dfrac{x}{4} & \text{para } 0 < x \leqq 2 \\[2ex] \dfrac{4}{x^3} & \text{para } x > 2 \\[2ex] 0 & \text{en cualquier otra parte} \end{cases}$$

Con respecto al inciso (b) del ejercicio 4.75, encuentre la longitud esperada de una de estas conversaciones telefónicas que ha durado al menos 1 minuto.

REFERENCIAS

En los textos más avanzados de estadística matemática listados al final del capítulo 3 se puede encontrar información adicional sobre el material de este capítulo.

CAPÍTULO

5

Distribuciones de probabilidad especiales

5.1 INTRODUCCIÓN

En este capítulo estudiaremos algunas de las distribuciones de probabilidad que figuran de manera más prominente en la teoría y las aplicaciones estadísticas. También estudiaremos sus **parámetros**; esto es, las cantidades que son constantes para distribuciones particulares pero que pueden tomar diferentes valores para diferentes miembros de las familias de distribuciones de la misma clase. Los parámetros más comunes son los momentos inferiores, principalmente μ y σ^2, y como vimos en el capítulo 4, hay esencialmente dos maneras en que se pueden obtener: podemos evaluar las sumas necesarias directamente o podemos trabajar con funciones generatrices de momentos. Aunque parecería lógico usar en cada caso cualquier método que fuera más simple, algunas veces usaremos ambos. En algunos casos esto se hará porque los resultados se necesitan más adelante; en otros meramente servirán para dar al lector experiencia en la aplicación de las técnicas matemáticas correspondientes. También, para mantener la extensión de este capítulo dentro de lo razonable, muchos de los detalles se dejan como ejercicios.

5.2 LA DISTRIBUCIÓN UNIFORME DISCRETA

Si una variable aleatoria discreta puede asumir k valores diferentes con igual probabilidad, decimos que tiene una **distribución uniforme discreta**; simbólicamente, DEFINICIÓN 5.1 Una variable aleatoria X tiene una **distribución uniforme discreta** y se conoce como una variable aleatoria uniforme discreta si y sólo si su distribución de probabilidad está dada por

$$f(x) = \frac{1}{k} \qquad \text{para } x = x_1, x_2, \ldots, x_k$$

donde $x_i \neq x_j$ cuando $i \neq j$.

De acuerdo a las definiciones 4.2 y 4.4, la media y la varianza de esta distribución son $\mu = \sum_{i=1}^{k} x_i \cdot \frac{1}{k}$ y $\sigma^2 = \sum_{i=1}^{k} (x_i - \mu)^2 \cdot \frac{1}{k}$.

En el caso especial donde $x_i = i$, la distribución uniforme discreta se vuelve $f(x) = \frac{1}{k}$ para $x = 1, 2, \ldots, k$, y en esta forma se aplica al número de puntos que lanzamos con un dado balanceado. La media y la varianza de esta distribución discreta uniforme y su función generatriz de momentos se tratan en los ejercicios 5.1 y 5.2.

5.3 LA DISTRIBUCIÓN DE BERNOULLI

Si un experimento tiene dos resultados posibles, "éxito" y "fracaso", y sus probabilidades son, respectivamente, θ y $1 - \theta$, entonces el número de éxitos, 0 o 1, tiene una **distribución de Bernoulli**; simbólicamente,

DEFINICIÓN 5.2 Una variable aleatoria X tiene una **distribución de Bernoulli** y se conoce como una variable aleatoria de Bernoulli si y sólo si su distribución de probabilidad está dada por

$$f(x; \theta) = \theta^x (1 - \theta)^{1-x} \qquad \text{para } x = 0, 1$$

Así, $f(0; \theta) = 1 - \theta$ y $f(1; \theta) = \theta$ se combinan en una sola fórmula. Observe que usamos la notación $f(x; \theta)$ para indicar explícitamente que la distribución de Bernoulli tiene el parámetro único θ. Puesto que la distribución de Bernoulli es un caso especial de la distribución de la sección 5.4, no la examinaremos aquí en detalle.

En relación con la distribución de Bernoulli, un éxito puede ser sacar cara con una moneda equilibrada, puede ser pescar una pulmonía, puede ser aprobar (o reprobar) un examen, y puede ser perder una carrera. Esta incongruencia es un remanente de los días cuando la teoría de probabilidad se aplicaba sólo a los juegos de azar (y el fracaso de un jugador era el éxito del otro). También por esta razón, nos referimos a un experimento al cual se aplica la distribución de Bernoulli como un **ensayo de Bernoulli**, o simplemente un **ensayo**, y a la serie de esos experimentos como **ensayos repetidos**.

5.4 LA DISTRIBUCIÓN BINOMIAL

Los ensayos repetidos juegan un papel muy importante en probabilidad y estadística, especialmente cuando el número de ensayos es fijo, el parámetro θ (la probabilidad de un éxito) es el mismo para cada ensayo, y los ensayos son todos independientes. Como veremos, surgen varias variables aleatorias en relación con ensayos repetidos. La que estudiaremos trata del número total de éxitos; en la sección 5.5 se darán otras.

La teoría que examinaremos en esta sección tiene muchas aplicaciones; por ejemplo, se aplica si queremos saber la probabilidad de sacar 5 caras en 12 lanzamientos de una moneda, la probabilidad de que 7 de 10 personas se recuperarán de una enfermedad tropical o la probabilidad de que 35 de 80 personas responderán a una solicitud por correo. Sin embargo, éste es el caso sólo si cada una de las 10 personas tiene la misma oportunidad de recuperarse de la enfermedad y su recuperación es independiente (digamos, que las tratan doctores diferentes en diferentes hospitales), y si la probabilidad de obtener una respuesta a la solicitud por correo es la misma para cada una de las 80 personas y hay independencia (digamos, no hay dos de ellas que pertenezcan al mismo hogar).

Para derivar una fórmula para la probabilidad de obtener "x éxitos en n ensayos" bajo las condiciones enunciadas, observe que la probabilidad de obtener x éxitos y $n - x$ fracasos en un *orden específico* es $\theta^x(1 - \theta)^{n-x}$. Hay un factor θ por cada éxito, un factor $1 - \theta$ por cada fracaso, y los x factores θ y $1 - \theta$ factores $n - x$ se multiplican todos entre sí en virtud de la suposición de independencia. Puesto que esta probabilidad se aplica a cualquier serie de n ensayos en la cual hay x éxitos y $n - x$ fracasos, sólo tenemos que contar cuántas series hay de esta clase y entonces multiplicar $\theta^x(1 - \theta)^{n-x}$ por ese número. Claramente, el número de maneras en que podemos seleccionar x ensayos en los que un éxito es $\binom{n}{x}$, y se sigue que la probabilidad deseada para "x éxitos en n ensayos" es $\binom{n}{x}\theta^x(1 - \theta)^{n-x}$.

> DEFINICIÓN 5.3 Una variable aleatoria X tiene una **distribución binomial** y se conoce como una variable aleatoria binomial si y sólo si su distribución de probabilidad está dada por
>
> $$b(x; n, \theta) = \binom{n}{x}\theta^x(1 - \theta)^{n-x} \qquad \text{para } x = 0, 1, 2, \ldots, n$$

Así, el número de éxitos en n ensayos es una variable aleatoria que tiene una distribución binomial con los parámetros n y θ. El nombre "distribución binomial" se deriva del hecho que los valores de $b(x; n, \theta)$ para $x = 0, 1, 2, \ldots, n$ son los términos sucesivos de la expansión binomial de $[(1-\theta)+\theta]^n$; esto demuestra también que la suma de las probabilidades es igual a 1, como debe ser.

EJEMPLO 5.1

Encuentre la probabilidad de sacar 5 caras y 7 cruces en 12 lanzamientos de una moneda equilibrada.

Solución

Al sustituir $x = 5$, $n = 12$ y $\theta = \frac{1}{2}$ en la fórmula de la distribución binomial, obtenemos

$$b\left(5; 12, \frac{1}{2}\right) = \binom{12}{5}\left(\frac{1}{2}\right)^5\left(1-\frac{1}{2}\right)^{12-5}$$

y, al buscar el valor de $\binom{12}{5}$ en la tabla VII, encontramos que el resultado es $792(\frac{1}{2})^{12}$ o aproximadamente 0.19. ▲

EJEMPLO 5.2

Encuentre la probabilidad de que siete de 10 personas se recuperarán de una enfermedad tropical si podemos suponer independencia y la probabilidad de que cualquiera de ellos se recuperará de la enfermedad es 0.80.

Solución

Al sustituir $x = 7$, $n = 10$ y $\theta = 0.80$ en la fórmula para la distribución binomial, obtenemos

$$b(7; 10, 0.80) = \binom{10}{7}(0.80)^7(1-0.80)^{10-7}$$

y, al buscar el valor de $\binom{10}{7}$ en la tabla VII, encontramos que el resultado es $120(0.80)^7(0.20)^3$ o aproximadamente 0.20. ▲

Si intentáramos calcular la tercera probabilidad pedida en la página 169, la tocante a las respuestas de una solicitud por correo, al sustituir $x = 35$, $n = 80$ y, digamos, $\theta = 0.15$, en la fórmula de la distribución binomial, encontraríamos que esto requiere una cantidad prohibitiva de trabajo. En la práctica real, rara vez se calculan directamente las probabilidades binomiales ya que hay extensas tabulaciones para diversos valores de θ y n, y existe una abundancia de software de computadoras que da las probabilidades binomiales así como las probabilidades acumuladas correspondientes

$$B(x; n, \theta) = \sum_{k=0}^{x} b(k; n, \theta)$$

con sólo dar instrucciones sencillas. En la figura 5.1 se muestra un ejemplo de una impresión así (con una notación algo diferente).

```
MTB > BINOMIAL N=10  P=0.63

  PROBABILIDADES BINOMIALES PARA N = 10 y P = .630000

     K              P(X = K)              P(X MENOR O = K)
     0               .0000                  .0000
     1               .0008                  .0009
     2               .0063                  .0071
     3               .0285                  .0356
     4               .0849                  .1205
     5               .1734                  .2939
     6               .2461                  .5400
     7               .2394                  .7794
     8               .1529                  .9323
     9               .0578                  .9902
    10               .0098                 1.0000
```

Figura 5.1 Impresión de computadora para probabilidades binomiales para $n = 10$ y $\theta = 0.63$.

Anteriormente, se usó ampliamente la tabla del National Bureau of Standards y el libro de H. G. Roming; se encuentran en la lista de las referencias al final de este capítulo. También, la tabla I da los valores de $b(x; n, \theta)$ con cuatro cifras decimales para $n = 1$ a $n = 20$ y $\theta = 0.05, 0.10, 0.15, \ldots, 0.45, 0.50$. Para usar esta tabla cuando θ es mayor que 0.50, nos referimos a la identidad

TEOREMA 5.1

$$b(x; n, \theta) = b(n - x; n, 1 - \theta)$$

cuya demostración se pedirá al lector en el inciso (a) del ejercicio 5.5. Por ejemplo, para encontrar $b(11; 18, 0.70)$, buscamos $b(7; 18, 0.30)$ y obtenemos 0.1376. También, hay varias maneras mediante las cuales se pueden aproximar las distribuciones binomiales cuando n es grande; en la sección 5.7 se menciona una de éstas y otra en la sección 6.6.

Hallemos ahora fórmulas para la media y la varianza de la distribución binomial.

TEOREMA 5.2 La media y la varianza de la distribución binomial son

$$\mu = n\theta \quad \text{y} \quad \sigma^2 = n\theta(1 - \theta)$$

Demostración

$$\begin{aligned}\mu &= \sum_{x=0}^{n} x \cdot \binom{n}{x} \theta^x (1-\theta)^{n-x} \\ &= \sum_{x=1}^{n} \frac{n!}{(x-1)!(n-x)!} \theta^x (1-\theta)^{n-x}\end{aligned}$$

donde omitimos el término correspondiente a $x = 0$, que es 0, y cancelamos la x contra el primer factor de $x! = x(x-1)!$ en el denominador de $\binom{n}{x}$. Entonces, sacamos el factor n en $n! = n(n-1)!$ y un factor θ, obtenemos

$$\mu = n\theta \cdot \sum_{x=1}^{n} \binom{n-1}{x-1} \theta^{x-1} (1-\theta)^{n-x}$$

y, al hacer $y = x - 1$ y $m = n - 1$, esto se vuelve

$$\mu = n\theta \cdot \sum_{y=0}^{m} \binom{m}{y} \theta^y (1-\theta)^{m-y} = n\theta$$

puesto que la última suma es la suma de todos los valores de una distribución binomial con parámetros m y θ, y por tanto es igual a 1.

Para encontrar expresiones para μ_2' y σ^2, usemos el hecho que $E(X^2) = E[X(X-1)] + E(X)$ y primero evaluamos $E[X(X-1)]$. Repetimos para todo propósito práctico los pasos antes usados y obtenemos así

$$\begin{aligned}E[X(X-1)] &= \sum_{x=0}^{n} x(x-1) \binom{n}{x} \theta^x (1-\theta)^{n-x} \\ &= \sum_{x=2}^{n} \frac{n!}{(x-2)!(n-x)!} \theta^x (1-\theta)^{n-x} \\ &= n(n-1)\theta^2 \cdot \sum_{x=2}^{n} \binom{n-2}{x-2} \theta^{x-2} (1-\theta)^{n-x}\end{aligned}$$

y, al hacer $y = x - 2$ y $m = n - 2$, esto se vuelve

$$\begin{aligned}E[X(X-1)] &= n(n-1)\theta^2 \cdot \sum_{y=0}^{m} \binom{m}{y} \theta^y (1-\theta)^{m-y} \\ &= n(n-1)\theta^2\end{aligned}$$

Por consiguiente,

$$\mu_2' = E[X(X-1)] + E(X) = n(n-1)\theta^2 + n\theta$$

y, finalmente,

$$\begin{aligned}\sigma^2 &= \mu_2' - \mu^2 \\ &= n(n-1)\theta^2 + n\theta - n^2\theta^2 \\ &= n\theta(1-\theta) \qquad \blacktriangledown\end{aligned}$$

En el ejercicio 5.6 se sugiere una demostración alternativa de este teorema que requiere mucho menos detalle algebraico.

No nos debe causar sorpresa que el producto $n\theta$. nos dé la media de la distribución binomial. Después de todo, si lanzamos una moneda equilibrada 200 veces, esperamos (en el sentido de esperanza matemática) $200 \cdot \frac{1}{2} = 100$ caras y 100 cruces; en forma similar, si un dado equilibrado se tira 240 veces, esperamos $240 \cdot \frac{1}{6} = 40$ seises, y si la probabilidad es 0.80 de que una persona de compras en una tienda departamental haga una compra, debemos esperar que $400(0.80) = 320$ de 400 personas de compras en la tienda departamental hagan una compra.

La fórmula de la varianza de la distribución binomial, al ser una medida de la variación, tiene muchas aplicaciones importantes; pero, para subrayar su importancia, consideremos la variable aleatoria $Y = \frac{X}{n}$, donde X es una variable aleatoria que tiene una distribución binomial con los parámetros n y θ. Esta variable aleatoria es la proporción de éxitos en n ensayos, y en el ejercicio 5.6 se le pedirá al lector que pruebe el siguiente resultado.

TEOREMA 5.3 Si X tiene una distribución binomial con los parámetros n y θ y $Y = \frac{X}{n}$, entonces

$$E(Y) = \theta \quad \text{y} \quad \sigma_Y^2 = \frac{\theta(1-\theta)}{n}$$

Ahora, si aplicamos el teorema de Chebyshev con $k\sigma = c$ (véase el ejercicio 4.40), podemos afirmar que *para cualquier constante positiva* **c** *la probabilidad es al menos*

$$1 - \frac{\theta(1-\theta)}{nc^2}$$

de que la proporción de éxitos en n ensayos caiga entre $\theta - c$ *y* $\theta + c$. Por tanto, *cuando* n S ∞, *la probabilidad se aproxima a 1 de que la proporción de éxitos diferirá de* θ *por menos que cualquier constante arbitraria* **c**. Este resultado se llama una **ley de grandes números**, y se debe observar que aplica a la proporción de éxitos, no a su número real. Es una falacia suponer que cuando n es grande el número de éxitos necesariamente debe estar próximo a $n\theta$.

Una fácil ilustración de esta ley de grandes números se puede obtener mediante una **simulación por computadora** de los lanzamientos repetidos de una moneda equilibrada. Esto se muestra en la figura 5.2 donde los 1 y los 0 denotan caras y cruces.

Al leer a lo ancho de renglones sucesivos, encontramos que entre los primeros cinco lanzamientos simulados hay 3 caras, entre los primeros diez hay 6 caras, entre los primeros 15 hay 8 caras, entre los primeros 20 hay 12 caras, entre los primeros 25 hay 14 caras, ..., y entre todos los cien hay 51 caras. Las proporciones correspondientes, cuya gráfica se muestra en la figura 5.3, son $\frac{3}{5} = 0.60$, $\frac{6}{10} = 0.60$, $\frac{8}{15} = 0.53$, $\frac{12}{20} = 0.60$,

```
MTB > BRANDOM 100 N=1  P=.5 C1

   100 EXPERIMENTOS BINOMIALES CON N = 1 Y P = .5000
     0.    0.    1. 1. 1.  1.   1.   0.   0.   1.
     1.    0.    0. 1. 0.  1.   1.   1.   0.   1.
     0.    0.    1. 0. 1.  1.   0.   1.   0.   0.
     1.    1.    0. 1. 0.  0.   1.   1.   1.   0.
     1.    0.    1. 0. 0.  0.   0.   1.   0.   0.
     1.    1.    0. 0. 0.  0.   0.   1.   0.   0.
     1.    1.    0. 0. 1.  1.   1.   0.   1.   1.
     1.    0.    1. 1. 0.  1.   1.   0.   0.   0.
     0.    0.    0. 1. 0.  0.   1.   0.   1.   1.
     1.    0.    1. 1. 1.  1.   0.   1.   0.   1.

   RESUMEN

   VALOR   FRECUENCIA
     0       49
     1       51
```

Figura 5.2 Simulación por computadora de 100 lanzamientos de una moneda equilibrada.

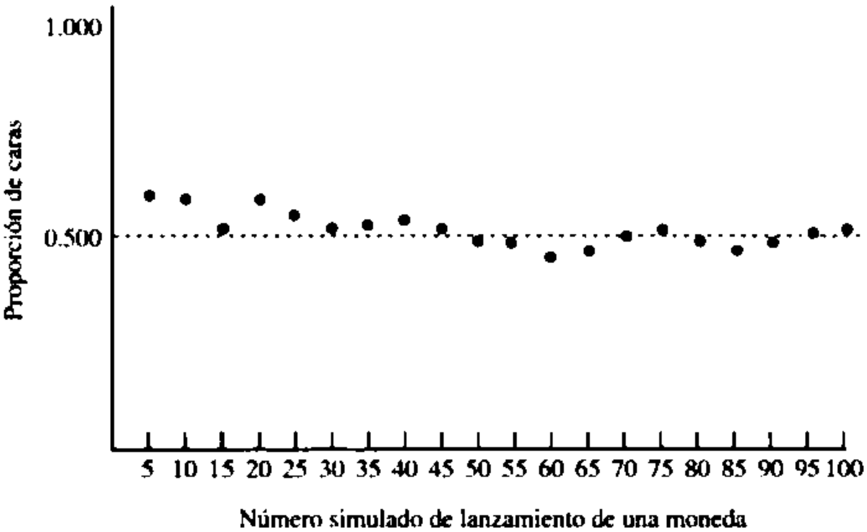

Figura 5.3 Gráfica que ilustra la ley de los grandes números.

$\frac{14}{25} = 0.56, \ldots$ y $\frac{51}{100} = 0.51$. Observe que la proporción de caras fluctúa pero se acerca cada vez más a 0.50, la probabilidad de cara en cada lanzamiento de la moneda.

Puesto que la función generatriz de momentos de la distribución binomial es fácil de obtener, encontrémosla y usémosla para verificar los resultados del teorema 5.2.

TEOREMA 5.4 La función generatriz de momentos de la distribución binomial está dada por

$$M_X(t) = [1 + \theta(e^t - 1)]^n$$

Demostración. Por las definiciones 4.6 y 5.3, obtenemos

$$M_X(t) = \sum_{x=0}^{n} e^{xt}\binom{n}{x}\theta^x(1-\theta)^{n-x}$$
$$= \sum_{x=0}^{n}\binom{n}{x}(\theta e^t)^x(1-\theta)^{n-x}$$

y por el teorema 1.9 esta suma se reconoce fácilmente como la expansión binomial de $[\theta e^t + (1-\theta)]^n = [1+\theta(e^t-1)]^n$. ▼

Si diferenciamos $M_X(t)$ dos veces con respecto a t, obtenemos

$$M'_X(t) = n\theta e^t[1+\theta(e^t-1)]^{n-1}$$
$$M''_X(t) = n\theta e^t[1+\theta(e^t-1)]^{n-1} + n(n-1)\theta^2e^{2t}[1+\theta(e^t-1)]^{n-2}$$
$$= n\theta e^t(1-\theta+n\theta e^t)[1+\theta(e^t-1)]^{n-2}$$

y, después de sustituir $t=0$, obtenemos $\mu'_1 = n\theta$ y $\mu'_2 = n\theta(1-\theta+n\theta)$. Así, $\mu = n\theta$ y $\sigma^2 = \mu'_2 - \mu^2 = n\theta(1-\theta+n\theta) - (n\theta)^2 = n\theta(1-\theta)$, lo cual concuerda con las fórmulas dadas en el teorema 5.2.

A partir del trabajo de esta sección puede parecer más fácil encontrar los momentos de la distribución binomial con la función generatriz de momentos que evaluarlas directamente, pero debe ser evidente que la diferenciación se vuelve bastante complicada si queremos determinar, digamos, μ'_3 o μ'_4. En realidad, existe todavía una forma más fácil de determinar los momentos de la distribución binomial; se basa en su **función factorial generatriz de momentos**, la cual se explica en el ejercicio 5.12.

EJERCICIOS

5.1 Si X tiene una distribución uniforme discreta $f(x) = \frac{1}{k}$ para $x = 1, 2, \ldots, k$, demuestre que

(a) su media es $\mu = \frac{k+1}{2}$;

(b) su varianza es $\sigma^2 = \frac{k^2-1}{12}$.

(*Sugerencia*: refiérase al apéndice A.)

5.2 Si X tiene una distribución uniforme discreta $f(x) = \frac{1}{k}$ para $x = 1, 2, \ldots, k$, demuestre que su función generatriz de momentos está dada por

$$M_X(t) = \frac{e^t(1-e^{kt})}{k(1-e^t)}$$

También encuentre la media de esta distribución al evaluar $\lim_{t\to 0} M'_X(t)$, y compare el resultado con el obtenido en el ejercicio 5.1.

5.3 En la sección 5.3 no estudiamos la distribución de Bernoulli con mucho detalle, porque puede considerarse como una distribución binomial con $n = 1$. Demuestre que para la distribución de Bernoulli, $\mu_r' = \theta$ para $r = 1, 2, 3, \ldots$, al

(a) evaluar la suma $\sum_{x=0}^{1} x^r \cdot f(x; \theta)$;

(b) hacer $n = 1$ en la función generatriz de momentos de la distribución binomial y al examinar su serie de Maclaurin.

5.4 Use el resultado del ejercicio 5.3 para demostrar que para la distribución de Bernoulli

(a) $\alpha_3 = \dfrac{1 - 2\theta}{\sqrt{\theta(1 - \theta)}}$, donde α_3 es la medida de la asimetría definida en el ejercicio 4.34;

(b) $\alpha_4 = \dfrac{1 - 3\theta(1 - \theta)}{\theta(1 - \theta)}$, donde α_4 es la medida de distribución puntiaguda definida en el ejercicio 4.35.

5.5 Verifique que

(a) $b(x; n, \theta) = b(n - x; n, 1 - \theta)$.

También demuestre que si $B(x; n, \theta) = \sum_{k=0}^{x} b(k; n, \theta)$ para $x = 0, 1, 2, \ldots, n$, entonces

(b) $b(x; n, \theta) = B(x; n, \theta) - B(x - 1; n, \theta)$;

(c) $b(x; n, \theta) = B(n - x; n, 1 - \theta) - B(n - x - 1; n, 1 - \theta)$;

(d) $B(x; n, \theta) = 1 - B(n - x - 1; n, 1 - \theta)$.

5.6 Una demostración alternativa del teorema 5.2 se puede basar en el hecho que si $X_1, X_2, \ldots$ y X_n son variables aleatorias independientes que tienen la misma distribución de Bernoulli con el parámetro θ, entonces $Y = X_1 + X_2 + \cdots + X_n$ es una variable aleatoria que tiene la distribución binomial con los parámetros n y θ.

(a) Verifique directamente (esto es, sin recurrir al hecho que la distribución de Bernoulli es un caso especial de la distribución binomial) que la media y la varianza de la distribución de Bernoulli son $\mu = \theta$ y $\sigma^2 = \theta(1 - \theta)$.

(b) Con base en el teorema 4.14 y su corolario en las páginas 158 y 159, demuestre que si $X_1, X_2, \ldots$ y X_n son variables aleatorias independientes que tienen la misma distribución de Bernoulli con el parámetro θ y $Y = X_1 + X_2 + \cdots + X_n$, entonces

$$E(Y) = n\theta \quad \text{y} \quad \operatorname{var}(Y) = n\theta(1 - \theta)$$

5.7 Demuestre el teorema 5.3.

5.8 Al calcular todos los valores de una distribución binomial, suele ser posible simplificar el trabajo al calcular primero $b(0; n, \theta)$ y después usar la fórmula recursiva

$$b(x + 1; n, \theta) = \frac{\theta(n - x)}{(x + 1)(1 - \theta)} \cdot b(x; n, \theta)$$

Verifique esta fórmula y úsela para calcular los valores de la distribución binomial con $n = 7$ y $\theta = 0.25$.

5.9 Use la fórmula recursiva del ejercicio 5.8 para demostrar que para $\theta = \frac{1}{2}$ la distribución binomial tiene

(a) un máximo en $x = \dfrac{n}{2}$ cuando n es par;

(b) máximos en $x = \dfrac{n-1}{2}$ y $x = \dfrac{n+1}{2}$ cuando n es impar.

5.10 Si X es una variable aleatoria binomial, ¿para qué valor de θ es la probabilidad de $b(x; n, \theta)$ un máximo?

5.11 En la demostración del teorema 5.2 determinamos la cantidad $E[X(X-1)]$, llamado el segundo **momento factorial**. En general, el $r$ésimo momento factorial de X está dado por

$$\mu'_{(r)} = E[X(X-1)(X-2)\cdot \ldots \cdot (X-r+1)]$$

Exprese μ'_2, μ'_3 y μ'_4 en términos de los momentos factoriales.

5.12 La **función generatriz de momentos factoriales** de una variable aleatoria discreta X está dada por

$$F_X(t) = E(t^X) = \sum_x t^x \cdot f(x)$$

Demuestre que la $r$ésima derivada de $F_X(t)$ con respecto a t en $t = 1$ es $\mu'_{(r)}$, el $r$ésimo momento factorial definido en el ejercicio 5.11.

5.13 Con respecto al ejercicio 5.12, encuentre la función generatriz de momentos factoriales de

(a) la distribución de Bernoulli y demuestre que $\mu'_{(1)} = \theta$ y $\mu'_{(r)} = 0$ para $r > 1$;

(b) la distribución binomial y úsela para encontrar μ y σ^2.

5.14 Si hacemos $a = -\mu$ en la primera parte del teorema 4.10, donde μ es la media de X, obtenemos

$$M_Y(t) = M_{X-\mu}(t) = e^{-\mu t} \cdot M_X(t)$$

(a) Demuestre que la $r$ésima derivada de $M_{X-\mu}(t)$ con respecto a t en $t = 0$ da el $r$ésimo momento alrededor de la media de X.

(b) Encuentre esa función generatriz para los momentos alrededor de la media de la distribución binomial, y verifique que la segunda derivada en $t = 0$ es $n\theta(1-\theta)$.

5.15 Use el resultado del inciso (b) del ejercicio 5.14 para demostrar que para la distribución binomial

$$\alpha_3 = \frac{1 - 2\theta}{\sqrt{n\theta(1-\theta)}}$$

donde α_3 es la medida de la asimetría definida en el ejercicio 4.34. ¿Qué podemos concluir sobre la asimetría de la distribución binomial cuando:

(a) $\theta = \frac{1}{2}$;

(b) n es grande?

APLICACIONES

5.16 Un examen de selección múltiple consiste en ocho preguntas y tres respuestas a cada pregunta (de las cuales sólo una es correcta). Si un estudiante contesta cada pregunta tirando un dado balanceado y marca la primera respuesta si saca 1 o 2, la segunda respuesta si saca 3 o 4, y la tercera respuesta si saca 5 o 6, ¿cuál es la probabilidad de que tendrá exactamente cuatro respuestas correctas?

5.17 Un ingeniero en seguridad automotriz afirma que 1 de 10 accidentes automovilísticos son causados por fatiga del conductor. Use la fórmula para la distribución binomial y con redondeo a cuatro decimales, ¿cuál es la probabilidad de que al menos 3 de 5 accidentes automovilísticos sean causados por fatiga del conductor?

5.18 Si 40 por ciento de los ratones que se usan en un experimento se vuelven muy agresivos dentro de 1 minuto después de habérseles administrado un medicamento experimental, encuentre la probabilidad de que exactamente seis de 15 ratones a los que se les ha administrado el medicamento se volverán muy agresivos dentro de 1 minuto, use

(a) la fórmula de la distribución binomial;

(b) la tabla I

5.19 En cierta ciudad, la incompatibilidad se da como la razón legal en 70 por ciento de todos los casos de divorcio. Encuentre la probabilidad que cinco de los seis casos siguientes de divorcio registrados en esta ciudad darán la incompatibilidad como razón legal, use

(a) la fórmula para la distribución binomial;

(b) la tabla I.

5.20 Un científico social afirma que sólo 50 por ciento de los estudiantes de último año de preparatoria capaces de desarrollar trabajo universitario van realmente a la universidad. Suponiendo que esta afirmación es cierta, use la tabla I para encontrar las probabilidades de que entre 18 estudiantes del último año de preparatoria capaces de hacer trabajo universitario

(a) exactamente 10 irán a la universidad;

(b) al menos 10 irán a la universidad;

(c) cuando mucho ocho irán a la universidad.

5.21 Suponga que la probabilidad de que un automóvil robado en cierta ciudad del oeste se recupere es 0.63. Use la impresión de computadora de la figura 5.1 para encontrar la probabilidad de que al menos 8 de 10 carros robados en esta ciudad se recuperarán, use

(a) los valores en la columna P(X = K);

(b) los valores en la columna P(X LESS OR = K).

5.22 Con respecto al ejercicio 5.21, encuentre la probabilidad de que entre 10 carros robados en la ciudad dada se recuperará cualquier cantidad entre tres y cinco, use

(a) los valores en la columna P(X = K);

(b) los valores en la columna P(X LESS OR = K).

5.23 Con respecto al ejercicio 5.18, suponga que el porcentaje haya sido 42 en vez de 40. Use una tabla apropiada o una impresión de computadora de la distribución binomial con $n = 15$ y $\theta = 0.42$ para volver a trabajar con ambas partes de ese ejercicio.

5.24 Con respecto al ejercicio 5.20, suponga que el porcentaje haya sido 51 en vez de 50. Use una tabla apropiada o una impresión de computadora de la distribución binomial con $n = 18$ y $\theta = 0.51$ para rehacer las tres partes de ese ejercicio.

5.25 Al planear la operación de una nueva escuela, un miembro del consejo de la escuela afirma que cuatro de cada cinco profesores recién contratados permanecerán en la escuela por más de un año, mientras que otro miembro del consejo de la escuela afirma que sería correcto decir que tres de cada cinco. En el pasado, los dos miembros del consejo han sido casi igualmente confiables en sus predicciones, así que en ausencia de cualquier otra información le asignaríamos a sus juicios igual peso. Si uno o el otro tiene que tener la razón ¿qué probabilidades le asignaríamos a sus afirmaciones si se encontrara que 11 de 12 profesores recientemente contratados permanecieron en la escuela por más de un año?

5.26 (a) Para reducir la desviación estándar de la distribución binomial a la mitad, ¿qué cambios debemos hacer en el número de ensayos?

(b) Si n es multiplicado por el factor k en la distribución binomial que tiene parámetros n y p, ¿qué afirmación se puede hacer sobre la desviación estándar de la distribución resultante?

5.27 Un fabricante afirma que cuando mucho 5 por ciento de las veces un producto dado aguantará menos de 1,000 horas de operación antes de requerir servicio. De la línea de producción se seleccionaron al azar veinte productos y se probaron. Se encontró que tres de ellos requirieron servicio antes de 1,000 horas de operación. Comente la afirmación del fabricante.

5.28 (a) Use un programa de cómputo para calcular la probabilidad de tirar entre 14 y 18 "sietes" en 100 tiradas de un par de dados.

(b) ¿Le sorprendería si se tiraran más de 18 "sietes"? ¿Por qué?

5.29 (a) Use un programa de cómputo para calcular la probabilidad de que más de 12 de 80 llamadas telefónicas duren más de cinco minutos si se supone que 10 por ciento de esas llamadas duran ese tiempo.

(b) ¿Se puede usar este resultado como evidencia de que la suposición es razonable? ¿Por qué?

5.30 Use el teorema de Chebyshev y el teorema 5.3 para verificar que la probabilidad es al menos $\frac{35}{36}$ de que

(a) en 900 lanzamientos de una moneda equilibrada la proporción de caras estará entre 0.40 y 0.60;

(b) en 10,000 lanzamientos de una moneda equilibrada la proporción estará entre 0.47 y 0.53;

(c) en 1,000,000 de lanzamientos de una moneda equilibrada la proporción de caras estará entre 0.497 y 0.503.

Advierta que esto sirve para ilustrar la ley de los grandes números.

5.31 Usted puede captar el sentido de la ley de los grandes números de la página 173 lanzando monedas. Lance una moneda 100 veces y haga una gráfica de la proporción acumulada de caras después de cada cinco lanzamientos.

5.32 Registre los primeros 200 números que se encuentra en un periódico, empiece con la página 1 y proceda en cualquier forma sistemática y conveniente. Incluya también los números que aparecen en los anuncios. Para cada uno de estos números tome nota del dígito de extrema izquierda y registre la proporción de los 1, los 2, los 3,... y los 9 (observe que 0 no puede ser el dígito de extrema izquierda; en el número decimal 0.0074, el dígito de extrema izquierda es 7). Los resultados pueden parecer muy sorprendentes, pero la ley de los grandes números dada en la página 173 dice que debe estar estimando correctamente.

5.5 LAS DISTRIBUCIONES BINOMIAL NEGATIVA Y GEOMÉTRICA

En relación con ensayos de Bernoulli repetidos, algunas veces nos interesa el número del ensayo en el cual ocurre el *k*ésimo éxito. Por ejemplo, puede interesarnos la probabilidad de que el décimo niño expuesto a una enfermedad contagiosa será el tercero en contagiarse, la probabilidad de que la quinta persona en escuchar un rumor será la primera en creerlo o la probabilidad de sorprender a un ladrón por segunda vez en su octavo robo.

Si el *k*ésimo éxito va a ocurrir en el ensayo *x*ésimo, debe haber $k - 1$ éxitos en los primeros $x - 1$ ensayos, y la probabilidad para esto es

$$b(k-1; x-1, \theta) = \binom{x-1}{k-1} \theta^{k-1}(1-\theta)^{x-k}$$

La probabilidad de un éxito en el *x*ésimo ensayo es θ, y la probabilidad de que el *k*ésimo éxito ocurra en el ensayo *x*ésimo es, por consiguiente,

$$\theta \cdot b(k-1; x-1, \theta) = \binom{x-1}{k-1} \theta^{k}(1-\theta)^{x-k}$$

DEFINICIÓN 5.4 Una variable aleatoria X tiene una **distribución binomial negativa** y se conoce como una variable aleatoria binomial negativa si y sólo si

$$b^*(x; k, \theta) = \binom{x-1}{k-1} \theta^{k}(1-\theta)^{x-k}$$

para $x = k, k+1, k+2, \ldots$.

Así, el número de ensayos en que ocurre el *késimo* éxito es una variable aleatoria que tiene una distribución binomial negativa con los parámetros y θ. El nombre "distribución binomial negativa" se deriva del hecho que los valores $b^*(x; k, \theta)$ para $x = k, k+1, k+2, \ldots$ son los términos sucesivos de la expansión binomial de

$\left(\frac{1}{\theta} - \frac{1-\theta}{\theta}\right)^{-k}$.† En la literatura estadística, las distribuciones binomiales negativas también se conocen como **distribuciones de tiempo de espera binomiales** o como **distribuciones de Pascal**.

EJEMPLO 5.3

Si la probabilidad es 0.40 de que un niño expuesto a una enfermedad contagiosa la contraiga, ¿cuál es la probabilidad de que el décimo niño expuesto a la enfermedad será el tercero en contraerla?

Solución

Sustituimos $x = 10$, $k = 3$ y $\theta = 0.40$ en la fórmula para la distribución binomial negativa, y obtenemos

$$\begin{aligned} b^*(10; 3, 0.40) &= \binom{9}{2}(0.40)^3(0.60)^7 \\ &= 0.0645 \quad \blacktriangle \end{aligned}$$

Cuando hay una tabla de probabilidades binomiales disponible, generalmente se puede simplificar la determinación de las probabilidades binomiales negativas mediante la identidad

TEOREMA 5.5

$$b^*(x; k, \theta) = \frac{k}{x} \cdot b(k; x, \theta)$$

Se pedirá al lector que verifique este teorema en el ejercicio 5.35.

EJEMPLO 5.4

Use el teorema 5.5 y la tabla 1 para rehacer el ejemplo 5.3.

Solución

Sustituimos $x = 10$, $k = 3$ y $\theta = 0.40$ en la fórmula del teorema 5.5, y obtenemos

$$\begin{aligned} b^*(10; 3, 0.40) &= \frac{3}{10} \cdot b(3; 10, 0.40) \\ &= \frac{3}{10}(0.2150) \\ &= 0.0645 \quad \blacktriangle \end{aligned}$$

† Las expansiones binomiales con exponentes negativos se explican en el libro de W. Feller mencionado entre las referencias al final de capítulo 2.

Los momentos de la distribución binomial negativa se pueden obtener al proceder como en la demostración del teorema 5.2; para la media y la varianza obtenemos

> **TEOREMA 5.6** La media y la varianza de la distribución binomial negativa son
>
> $$\mu = \frac{k}{\theta} \quad \text{y} \quad \sigma^2 = \frac{k}{\theta}\left(\frac{1}{\theta} - 1\right)$$

como se pedirá al lector que verifique en el ejercicio 5.36.

Puesto que la distribución binomial negativa con $k = 1$ tiene muchas aplicaciones importantes, se le ha dado un nombre especial: **distribución geométrica**.

> **DEFINICIÓN 5.5** Una variable aleatoria X tiene una **distribución geométrica** y se le conoce como una variable aleatoria geométrica si y sólo si su distribución de probabilidad está dada por
>
> $$g(x; \theta) = \theta(1 - \theta)^{x-1} \qquad \text{para } x = 1, 2, 3, \ldots$$

EJEMPLO 5.5

Si la probabilidad es 0.75 de que el solicitante de una licencia de manejo pasará la prueba de manejo en un ensayo dado, ¿cuál es la probabilidad de que un solicitante finalmente pase la prueba en el cuarto ensayo?

Solución

Al sustituir $x = 4$ y $\theta = 0.75$ en la fórmula de la distribución geométrica, obtenemos

$$\begin{aligned} g(4; 0.75) &= 0.75(1 - 0.75)^{4-1} \\ &= 0.75(0.25)^3 \\ &= 0.0117 \end{aligned}$$

Por supuesto, este resultado se basa en la suposición de que las pruebas son independientes y aquí puede haber algunas dudas sobre su validez. ▲

5.6 LA DISTRIBUCIÓN HIPERGEOMÉTRICA

En el capítulo 2 usamos el muestreo con y sin reemplazo para ilustrar las reglas de multiplicación para eventos independientes y dependientes. Para obtener una fórmula análoga a la de la distribución binomial que sea válida para el muestreo sin reemplazo, en cuyo caso los ensayos no son independientes, consideremos un conjunto de N elementos de los cuales M se consideran como éxitos y los otros $N - M$ como fracasos. Así como en relación con la distribución binomial, estamos interesados en la probabilidad de obtener x éxitos en n ensayos, pero ahora estamos escogiendo, sin reemplazo, n de los N elementos contenidos en el conjunto.

Hay $\binom{M}{x}$ maneras de escoger x de los M éxitos y $\binom{N-M}{n-x}$ maneras de escoger $n - x$ de los $N - M$ fracasos, y, por tanto, $\binom{M}{x}\binom{N-M}{n-x}$ maneras de escoger x éxitos y $n - x$ fracasos. Puesto que hay $\binom{N}{n}$ maneras de escoger n de los N elementos en el conjunto, y supondremos que son todos igualmente posibles (que es lo que queremos decir cuando afirmamos que la selección es al azar), se sigue del teorema 2.2 que la probabilidad de "x éxitos en n ensayos" es

$$\binom{M}{x}\binom{N-M}{n-x}\bigg/\binom{N}{n}.$$

DEFINICIÓN 5.6 Una variable aleatoria X tiene una **distribución hipergeométrica** y se conoce como variable aleatoria hipergeométrica si y sólo si su distribución de probabilidad está dada por

$$h(x; n, N, M) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} \qquad \begin{array}{l} \text{para } x = 0, 1, 2, \ldots, n, \\ x \leqq M \text{ y } n - x \leqq N - M \end{array}$$

Así, para el muestreo sin reemplazos, el número de éxitos en n ensayos es una variable aleatoria que tiene una distribución hipergeométrica con los parámetros n, N y M.

EJEMPLO 5.6

Como parte de una encuesta de contaminación del aire, un inspector decide examinar las emisiones de seis de los 24 camiones de una compañía. Si cuatro de los camiones de la compañía emiten cantidades excesivas de contaminantes, ¿cuál es la probabilidad de que ninguno de ellos sea parte de la muestra del inspector?

Solución

Al sustituir $x = 0$, $n = 6$, $N = 24$ y $M = 4$ en la fórmula para la distribución hipergeométrica, obtenemos

$$h(0; 6, 24, 4) = \frac{\binom{4}{0}\binom{20}{6}}{\binom{24}{6}}$$

$$= 0.2880 \qquad \blacktriangle$$

El método por el cual encontramos la media y la varianza de la distribución hipergeométrica es muy similar al empleado en la demostración del teorema 5.2.

> **TEOREMA 5.7** La media y la varianza de la distribución hipergeométrica son
>
> $$\mu = \frac{nM}{N} \quad \text{y} \quad \sigma^2 = \frac{nM(N-M)(N-n)}{N^2(N-1)}$$

Demostración. Para determinar la media, evaluemos directamente la suma

$$\mu = \sum_{x=0}^{n} x \cdot \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

$$= \sum_{x=1}^{n} \frac{M!}{(x-1)!(M-x)!} \cdot \frac{\binom{N-M}{n-x}}{\binom{N}{n}}$$

donde omitimos el término correspondiente a $x = 0$, el cual es 0, y cancelamos la x contra el primer factor de $x! = x(x-1)!$ en el denominador de $\binom{M}{x}$. Entonces, sacamos como factor $M\Big/\binom{N}{n}$, y obtenemos

$$\mu = \frac{M}{\binom{N}{n}} \cdot \sum_{x=1}^{n} \binom{M-1}{x-1}\binom{N-M}{n-x}$$

y, al hacer $y = x - 1$ y $m = n - 1$, esto se convierte en

$$\mu = \frac{M}{\binom{N}{n}} \cdot \sum_{y=0}^{m} \binom{M-1}{y}\binom{N-M}{m-y}$$

Finalmente, usamos el teorema 1.12, y obtenemos

$$\mu = \frac{M}{\binom{N}{n}} \cdot \binom{N-1}{m} = \frac{M}{\binom{N}{n}} \cdot \binom{N-1}{n-1} = \frac{nM}{N}$$

A fin de obtener la fórmula para σ^2, procedemos como en la demostración del teorema 5.2, evaluamos primero $E[X(X-1)]$ y entonces usamos el hecho que $E(X^2) = E[X(X-1)] + E(X)$. Le dejamos al lector demostrar que

$$E[X(X-1)] = \frac{M(M-1)n(n-1)}{N(N-1)}$$

en el ejercicio 5.44, obtenemos así

$$\sigma^2 = \frac{M(M-1)n(n-1)}{N(N-1)} + \frac{nM}{N} - \left(\frac{nM}{N}\right)^2$$

$$= \frac{nM(N-M)(N-n)}{N^2(N-1)} \qquad \blacktriangledown$$

Puesto que la función generatriz de momentos de la distribución hipergeométrica es bastante complicada, no la trataremos en este libro. Sin embargo, se pueden encontrar los detalles en el libro de M. G. Kendall y A. Stuart que se enumera entre las referencias al final del capítulo 3.

Cuando N es grande y n es relativamente pequeña comparada con N (la regla empírica es que n no debe exceder de 5 por ciento de N), no hay mucha diferencia entre muestrear con reemplazo y muestrear sin reemplazo y se puede usar la fórmula para la distribución binomial con los parámetros n y $\theta = \frac{M}{N}$ para aproximar las probabilidades hipergeométricas.

EJEMPLO 5.7

Entre los 120 solicitantes para un trabajo, sólo 80 son realmente aptos. Si cinco de los solicitantes se seleccionan al azar para una entrevista más extensa, encuentre la probabilidad de que sólo dos de los cinco serán aptos para el trabajo, para ello use

(a) la fórmula para la distribución hipergeométrica;
(b) la fórmula para la distribución binomial con $\theta = \frac{80}{120}$ como una aproximación.

Solución

(a) Al sustituir $x = 2$, $n = 5$, $N = 120$ y $M = 80$ en la fórmula de la distribución hipergeométrica, obtenemos

$$h(2; 5, 120, 80) = \frac{\binom{80}{2}\binom{40}{3}}{\binom{120}{5}}$$

$$= 0.164$$

redondeado a tres decimales;

(b) Al sustituir $x = 2$, $n = 5$ y $\theta = \frac{80}{120} = \frac{2}{3}$ en la fórmula para la distribución binomial, obtenemos

$$b\left(2; 5, \frac{2}{3}\right) = \binom{5}{2}\left(\frac{2}{3}\right)^2\left(1 - \frac{2}{3}\right)^3$$
$$= 0.165$$

redondeado a tres decimales. Como se puede ver a partir de estos resultados la aproximación es muy cercana. ▲

5.7 LA DISTRIBUCIÓN DE POISSON

Cuando n es grande, el cálculo de las probabilidades binomiales con la fórmula de la definición 5.3 implica una cantidad prohibitiva de trabajo. Por ejemplo, para calcular la probabilidad de que 18 de 3,000 personas, que ven un desfile en un día muy caluroso de verano, sufrirán de insolación, primero debemos determinar $\binom{3{,}000}{18}$, y si la probabilidad es 0.005 de que cualquiera de las 3,000 personas que ven el desfilen sufrirán de insolación, también tenemos que calcular el valor de $(0.005)^{18}(0.995)^{2{,}982}$.

En esta sección presentaremos una distribución de probabilidad que se puede usar para aproximar probabilidades binomiales de esta clase. Específicamente, investigaremos la forma límite de la distribución binomial cuando $n \to \infty$, $\theta \to 0$, mientras $n\theta$ permanece constante. Sea esta constante λ, esto es, $n\theta = \lambda$ y, por tanto, $\theta = \frac{\lambda}{n}$, podemos escribir

$$b(x; n, \theta) = \binom{n}{x}\left(\frac{\lambda}{n}\right)^x\left(1 - \frac{\lambda}{n}\right)^{n-x}$$
$$= \frac{n(n-1)(n-2)\cdot \ldots \cdot(n-x+1)}{x!}\left(\frac{\lambda}{n}\right)^x\left(1 - \frac{\lambda}{n}\right)^{n-x}$$

Entonces, si dividimos uno de los x factores n en $\left(\frac{\lambda}{n}\right)^x$ cada factor del producto $n(n-1)$ $(n-2)\cdot \ldots \cdot(n-x+1)$ y escribimos

$$\left(1 - \frac{\lambda}{n}\right)^{n-x} \quad \text{como} \quad \left[\left(1 - \frac{\lambda}{n}\right)^{-n/\lambda}\right]^{-\lambda}\left(1 - \frac{\lambda}{n}\right)^{-x}$$

obtenemos

$$\frac{1\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\cdot \ldots \cdot\left(1 - \frac{x-1}{n}\right)}{x!}(\lambda)^x\left[\left(1 - \frac{\lambda}{n}\right)^{-n/\lambda}\right]^{-\lambda}\left(1 - \frac{\lambda}{n}\right)^{-x}$$

Finalmente, si hacemos $n \to \infty$ mientras x y λ permanecen fijas, encontramos que

$$1\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdot\ldots\cdot\left(1-\frac{x-1}{n}\right)\to 1$$

$$\left(1-\frac{\lambda}{n}\right)^{-x}\to 1$$

$$\left(1-\frac{\lambda}{n}\right)^{-n/\lambda}\to e$$

y, por tanto, que la distribución límite se vuelve

$$p(x;\lambda)=\frac{\lambda^x e^{-\lambda}}{x!}\qquad \text{para } x=0,1,2,\ldots$$

DEFINICIÓN 5.7 Una variable aleatoria X tiene una **distribución de Poisson** y se conoce como una variable aleatoria de Poisson si y sólo si su distribución de probabilidad está dada por

$$p(x;\lambda)=\frac{\lambda^x e^{-\lambda}}{x!}\qquad \text{para } x=0,1,2,\ldots$$

Así, en el límite cuando $n \to \infty$, $\theta \to 0$, y $n\theta = \lambda$ permanece constante, el número de éxitos es una variable aleatoria que tiene una distribución de Poisson con el parámetro λ. Esta distribución se llama así en honor al matemático francés Simeon Poisson (1781 - 1840). En general, la distribución de Poisson brindará una buena aproximación a las probabilidades binomiales cuando $n \geqq 20$ y $\theta \leqq 0.05$. Cuando $n \geqq 100$ y $n\theta < 10$, la aproximación generalmente será excelente.

Para tener una idea sobre la cercanía de la aproximación de Poisson a la distribución binomial, considere la impresión de computadora de la figura 5.4, que muestra, una arriba de la otra, la distribución binomial con $n = 150$ y $\theta = 0.05$ y la distribución de Poisson con $\lambda = 150(0.05) = 7.5$.

EJEMPLO 5.8

Use la figura 5.4 para determinar el valor de x (desde 5 hasta 15) para el cual el error es el más grande cuando usamos la distribución de Poisson con $\lambda = 7.5$ para aproximar la distribución binomial con $n = 150$ y $\theta = 0.05$.

Solución

Calculamos las diferencias correspondientes a $x = 5, x = 6, \ldots, x = 15$, obtenemos 0.0006, -0.0017, -0.0034, -0.0037, -0.0027, -0.0011, 0.0003, 0.0011, 0.0013, 0.0011 y 0.0008. Así, el máximo error (numéricamente) es -0.0037, y corresponde a $x = 8$. ▲

```
MTB > BINOMIAL N=150  P=0.05

  PROBABILIDADES BINOMIALES MENOR N = 1.50 Y P = .050000

    K      P(X = K)  P(X PARA O = K)
    0       .0005        .0005
    1       .0036        .0041
    2       .0141        .0182
    3       .0366        .0548
    4       .0708        .1256
    5       .1088        .2344
    6       .1384        .3729
    7       .1499        .5228
    8       .1410        .6638
    9       .1171        .7809
   10       .0869        .8678
   11       .0582        .9260
   12       .0355        .9615
   13       .0198        .9813
   14       .0102        .9915
   15       .0049        .9964
   16       .0022        .9986
   17       .0009        .9995
   18       .0003        .9998
   19       .0001        .9999

MTB > POISSON MU=7.5

  PROBABILIDADES DE POISSON PARA MEDIA = 7.500

    K      P(X = K)  P(X MENOR O = K)
    0       .0006        .0006
    1       .0041        .0047
    2       .0156        .0203
    3       .0389        .0591
    4       .0729        .1321
    5       .1094        .2414
    6       .1367        .3782
    7       .1465        .5246
    8       .1373        .6620
    9       .1144        .7764
   10       .0858        .8622
   11       .0585        .9208
   12       .0366        .9573
   13       .0211        .9784
   14       .0113        .9897
   15       .0057        .9954
   16       .0026        .9980
   17       .0012        .9992
   18       .0005        .9997
   19       .0002        .9999
   20       .0001       1.0000
```

Figura 5.4 Impresión de computadora de la distribución binomial con $n = 150$ y $\theta = 0.05$ y la distribución de Poisson con $\lambda = 7.5$.

El siguiente ejemplo ilustra la aproximación de Poisson a la distribución binomial.

EJEMPLO 5.9

Si 2 por ciento de los libros encuadernados en cierto taller tiene encuadernación defectuosa, use la aproximación de Poisson a la distribución binomial para determinar la probabilidad de que cinco de 400 libros encuadernados en este taller tendrán encuadernaciones defectuosas.

Solución

Al sustituir $x = 5$, $\lambda = 400(0.02) = 8$ y $e^{-8} = 0.00034$ (de la tabla VIII) en la fórmula de la definición 5.7, obtenemos

$$p(5; 8) = \frac{8^5 \cdot e^{-8}}{5!} = \frac{(32{,}768)(0.00034)}{120} = 0.093 \qquad \blacktriangle$$

En la práctica real, rara vez se obtienen las probabilidades de Poisson por sustitución directa en la fórmula de la definición 5.7. Algunas veces nos referimos a las tablas de probabilidades de Poisson, como la tabla II, o a tablas más extensas en los manuales de tablas estadísticas, pero más a menudo, hoy en día, nos referimos a programas de computadora apropiados. El uso de tablas o computadoras es de especial importancia cuando nos interesan probabilidades relacionadas con varios valores de x.

EJEMPLO 5.10

Los registros muestran que la probabilidad es 0.00005 de que a un automóvil se le reviente un neumático mientras cruza cierto puente. Use la distribución de Poisson para aproximar las probabilidades binomiales que, de 10,000 autos que cruzan este puente,

(a) exactamente dos tendrán un neumático reventado;

(b) cuando mucho dos tendrán un neumático reventado.

Solución

(a) Al consultar la tabla II, encontramos que para $x = 2$ y $\lambda = 10{,}000(0.00005) = 0.5$, la probabilidad de Poisson es 0.0758.

(b) Al consultar la tabla II, encontramos que para $x = 0, 1$ y 2, y $\lambda = 0.5$, las probabilidades de Poisson son 0.6065, 0.3033 y 0.0758. Así, la probabilidad de que cuando mucho dos de los 10,000 autos que cruzan el puente tendrán un neumático reventado es

$$0.6065 + 0.3033 + 0.0758 = 0.9856 \qquad \blacktriangle$$

EJEMPLO 5.11

Use la figura 5.5 para rehacer el ejemplo anterior.

Solución

(a) Leemos el valor para $K = 2$ en la columna P(X = K) obtenemos 0.0758.

(b) En este caso podemos sumar los valores para $K = 0$, $K = 1$ y $K = 2$ en la columna P(X = K) o podemos leer el valor para $K = 2$ en la columna P(X LESS OR = K) y obtener 0.9856. ▲

```
MTB > POISSON MU=.5
  PROBABILIDADES DE POISSON PARA MEDIA = .500

     K      P(X = K)  P(X MENOR O = K)
     0       .6065             .6065
     1       .3033             .9098
     2       .0758             .9856
     3       .0126             .9982
     4       .0016             .9998
     5       .0002            1.0000
```

Figura 5.5 Impresión de computadora de la distribución de Poisson con $\lambda = 0.5$.

Al haber derivado la distribución de Poisson como una forma límite de la distribución binomial, podemos obtener fórmulas para su media y su varianza al aplicar las mismas condiciones límite ($n \to \infty$, $\theta \to 0$ y $n\theta = \lambda$ permanece constante) a la media y la varianza de la distribución binomial. Para la media obtenemos $\mu = n\theta = \lambda$ y para la varianza obtenemos $\sigma^2 = n\theta(1 - \theta) = \lambda(1 - \theta)$, la cual se aproxima a λ cuando $\theta \to 0$.

TEOREMA 5.8 La media y la varianza de la distribución de Poisson están dadas por

$$\mu = \lambda \quad \text{y} \quad \sigma^2 = \lambda$$

Estos resultados también se pueden obtener al evaluar directamente las sumas necesarias (véase el ejercicio 5.50) o al trabajar con la función generatriz de momentos dada en el siguiente teorema

TEOREMA 5.9 La función generatriz de momentos de la distribución de Poisson está dada por

$$M_X(t) = e^{\lambda(e^t - 1)}$$

Demostración. Por las definiciones 4.6 y 5.7,

$$M_X(t) = \sum_{x=0}^{\infty} e^{xt} \cdot \frac{\lambda^x e^{-\lambda}}{x!} = e^{-\lambda} \cdot \sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!}$$

donde $\sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!}$ se puede reconocer como la serie de Maclaurin de e^z con $z = \lambda e^t$. Así,

$$M_X(t) = e^{-\lambda} \cdot e^{\lambda e^t} = e^{\lambda(e^t - 1)} \qquad \blacktriangledown$$

Entonces, si diferenciamos $M_X(t)$ dos veces con respecto a t, obtenemos

$$M'_X(t) = \lambda e^t e^{\lambda(e^t - 1)}$$

$$M''_X(t) = \lambda e^t e^{\lambda(e^t-1)} + \lambda^2 e^{2t} e^{\lambda(e^t-1)}$$

de manera que $\mu'_1 = M'_X(0) = \lambda$ y $\mu'_2 = M''_X(0) = \lambda + \lambda^2$. Así, $\mu = \lambda$ y $\sigma^2 = \mu'_2 - \mu^2 = (\lambda + \lambda^2) - \lambda^2 = \lambda$, lo cual concuerda con el teorema 5.8.

Aunque la distribución de Poisson se ha derivado como una forma límite de la distribución binomial, tiene muchas aplicaciones que no tienen relación directa con distribuciones binomiales. Por ejemplo, la distribución de Poisson puede servir como un modelo para el número de éxitos que ocurren durante un intervalo de tiempo dado o en una región específica cuando (1) son independientes los números de éxitos que ocurren en intervalos de tiempo o regiones que no se traslapan, (2) la probabilidad de que ocurra un solo éxito en un intervalo de tiempo muy corto o en una región muy pequeña es proporcional a la longitud del intervalo de tiempo o al tamaño de la región, y (3) la probabilidad de que más de un éxito ocurra en tal intervalo corto de tiempo o que caiga en tal región pequeña es insignificante. De ahí, una distribución de Poisson podría describir el número de llamadas telefónicas recibidas por hora en una oficina, el número de errores de mecanografía por página o el número de bacterias en un cultivo dado cuando se conoce el número promedio de éxitos, λ, para el intervalo de tiempo o la región específica dados.

EJEMPLO 5.12

El número de camiones que llegan en un día cualquiera a un paradero de camiones en cierta ciudad se conoce que es 12. ¿Cuál es la probabilidad de que en un día dado lleguen menos de 9 camiones a ese paradero?

Solución

Sea X el número de camiones que llegan en un día dado. Entonces, usamos la tabla II con $\lambda = 12$, y obtenemos

$$P(X < 9) = \sum_{x=0}^{8} p(x; 12) = 0.1550 \qquad \blacktriangle$$

Si, en una situación donde son válidas las condiciones anteriores, los éxitos ocurren a una tasa media de α por *unidad* de tiempo o por *unidad* de región, entonces el número de éxitos en un intervalo de t unidades de tiempo o t unidades de la región especificada es una variable aleatoria de Poisson con la media $\lambda = \alpha t$ (véase el ejercicio 5.48). Por consiguiente, el número de éxitos, X, en un intervalo de tiempo de longitud t unidades o una región de tamaño t unidades tiene la distribución de Poisson

$$p(x; \alpha t) = \frac{e^{-\alpha t}(\alpha t)^x}{x!} \qquad \text{para } x = 0, 1, 2, \ldots$$

EJEMPLO 5.13

Cierta clase de lámina de metal tiene, en promedio, cinco defectos por cada 10 pies cuadrados. Si suponemos una distribución de Poisson, ¿cuál es la probabilidad de que una lámina del metal de 15 pies cuadrados tendrá al menos 6 defectos?

Solución

Sea X el número de defectos en una lámina del metal de 15 pies cuadrados. Entonces, puesto que la unidad de área es 10 pies cuadrados, tenemos

$$\lambda = \alpha t = (5)(1.5) = 7.5$$

y

$$P(X \geqq 6) = 1 - P(X \leqq 5) = 1 - 0.2414 = 0.7586$$

de acuerdo a la impresión de computadora mostrada en la figura 5.4. ▲

EJERCICIOS

5.33 A veces se define la distribución binomial negativa en una forma diferente como la distribución del número de fracasos que precede el $k$ésimo éxito. Si el $k$ésimo éxito ocurre en el $x$ésimo ensayo, debe estar precedido por $x - k$ fracasos. Así, encuentre la distribución de $Y = X - k$, donde X tiene la distribución de la definición 5.4.

5.34 Con respecto al ejercicio 5.33, encuentre expresiones para μ_Y y σ_Y^2.

5.35 Demuestre el teorema 5.5.

5.36 Demuestre el teorema 5.6 al determinar primero $E(X)$ y $E[X(X+1)]$.

5.37 Demuestre que la función generatriz de momentos de la distribución geométrica está dada por

$$M_X(t) = \frac{\theta e^t}{1 - e^t(1 - \theta)}$$

5.38 Use la función generatriz de momentos derivada en el ejercicio 5.37 para mostrar que para la distribución geométrica, $\mu = \frac{1}{\theta}$ y $\sigma^2 = \frac{1 - \theta}{\theta^2}$.

5.39 Al diferenciar con respecto a θ las expresiones en ambos lados de la ecuación

$$\sum_{x=1}^{\infty} \theta(1 - \theta)^{x-1} = 1$$

demuestre que la media de la distribución geométrica está dada por $\mu = \frac{1}{\theta}$.

Entonces, al diferenciar otra vez con respecto a θ, demuestre que $\mu_2' = \frac{2-\theta}{\theta^2}$ y por tanto que $\sigma^2 = \frac{1-\theta}{\theta^2}$.

5.40 Si X es una variable aleatoria que tiene una distribución geométrica, demuestre que

$$P(X = x + n | X > n) = P(X = x)$$

5.41 Si la probabilidad es $f(x)$ de que un producto falle la $x$ésima vez que se usa, esto es, en el $x$ésimo ensayo, entonces su **tasa de falla** en el $x$ésimo ensayo es la probabilidad de que fallará en el $x$ésimo ensayo dado que no ha fallado en los primeros $x - 1$ ensayos; simbólicamente, está dada por

$$Z(x) = \frac{f(x)}{1 - F(x-1)}$$

donde $F(x)$ es el valor de la función de distribución correspondiente en x. Demuestre que si X es una variable aleatoria geométrica, su tasa de falla es constante e igual a θ.

5.42 Se presenta una variación de la distribución binomial cuando los n ensayos son todos independientes, pero la probabilidad de éxito en el $i$ésimo ensayo es θ_i, y estas probabilidades no son todas iguales. Si X es el número de éxitos obtenidos bajo estas condiciones en n ensayos, demuestre que

(a) $\mu_X = n\theta$, donde $\theta = \frac{1}{n} \cdot \sum_{i=1}^{n} \theta_i$;

(b) $\sigma_X^2 = n\theta(1-\theta) - n\sigma_\theta^2$, donde θ es como se define en el inciso (a) y $\sigma_\theta^2 = \frac{1}{n} \cdot \sum_{i=1}^{n} (\theta_i - \theta)^2$.

5.43 Cuando se calculan todos los valores de una distribución hipergeométrica, a menudo se puede simplificar el trabajo al calcular primero $h(0; n, N, M)$ y usar entonces la fórmula recursiva

$$h(x+1; n, N, M) = \frac{(n-x)(M-x)}{(x+1)(N-M-n+x+1)} \cdot h(x; n, N, M)$$

Verifique esta fórmula y úsela para calcular los valores de la distribución hipergeométrica con $n = 4$, $N = 9$ y $k = 5$.

5.44 Verifique la expresión dada para $E[X(X-1)]$ en la demostración del teorema 5.7.

5.45 Demuestre que si hacemos $\theta = \frac{M}{N}$ en el teorema 5.7, la media y la varianza de la distribución hipergeométrica se pueden escribir como $\mu = n\theta$ y $\sigma^2 = n\theta(1-\theta) \cdot \frac{N-n}{N-1}$. ¿Cómo se comparan estos resultados con el análisis de la página 185?

5.46 Cuando calculamos todos los valores de una distribución de Poisson, a menudo se puede simplificar el trabajo al calcular primero $p(0; \lambda)$ y usar después la fórmula recursiva

$$p(x+1; \lambda) = \frac{\lambda}{x+1} \cdot p(x; \lambda)$$

Verifique esta fórmula y úsela junto con $e^{-2} = 0.1353$ para verificar los valores dados en la tabla II para $\lambda = 2$.

5.47 Aproxime la probabilidad binomial $b(3; 100, 0.10)$ al usar

(a) la fórmula para la distribución binomial y logaritmos;

(b) la tabla II

5.48 Suponga que $f(x, t)$ es la probabilidad de obtener x éxitos durante un intervalo de tiempo de longitud t cuando (i) la probabilidad de un éxito durante un intervalo de tiempo muy pequeño de t a $t + \Delta t$ es $\alpha \cdot \Delta t$, (ii) la probabilidad de más de un éxito durante dicho intervalo de tiempo es insignificante, y (iii) la probabilidad de un éxito durante ese intervalo de tiempo no depende de lo que haya pasado antes del tiempo t.

(a) Demuestre que bajo estas condiciones

$$f(x, t + \Delta t) = f(x, t)[1 - \alpha \cdot \Delta t] + f(x - 1, t)\alpha \cdot \Delta t$$

y por tanto que

$$\frac{d[f(x, t)]}{dt} = \alpha[f(x - 1, t) - f(x, t)]$$

(b) Demuestre por sustitución directa que una solución a este sistema infinito de ecuaciones diferenciales (hay una para cada valor de x) está dada por la distribución de Poisson con $\lambda = \alpha t$.

5.49 Use repetidamente la integración por partes para demostrar que

$$\sum_{y=0}^{x} \frac{\lambda^y e^{-\lambda}}{y!} = \frac{1}{x!} \cdot \int_{\lambda}^{\infty} t^x e^{-t}\, dt$$

Este resultado es importante porque los valores de la función de distribución de una variable aleatoria de Poisson se pueden obtener así al consultar una tabla de funciones gamma incompletas.

5.50 Derive las fórmulas para la media y la varianza de la distribución de Poisson al evaluar primero $E(X)$ y $E[X(X - 1)]$.

5.51 Demuestre que si las condiciones límite $n \to \infty$, $\theta \to 0$, mientras $n\theta$ permanece constante, se aplican a la función generatriz de momentos de la distribución binomial, obtenemos la función generatriz de momentos de la distribución de Poisson. [*Sugerencia*: válgase del hecho que $\lim_{n\to\infty} \left(1 + \frac{z}{n}\right)^n = e^z$.]

5.52 Use el teorema 5.9 para mostrar que para la distribución de Poisson $\alpha_3 = \frac{1}{\sqrt{\lambda}}$, donde α_3 es la medida de la asimetría definida en el ejercicio 4.34.

5.53 Al diferenciar con respecto a λ las expresiones de ambos lados de la ecuación

$$\mu_r = \sum_{x=0}^{\infty} (x - \lambda)^r \cdot \frac{\lambda^x e^{-\lambda}}{x!}$$

derive la siguiente fórmula recursiva para los momentos alrededor de la media de la distribución de Poisson:

$$\mu_{r+1} = \lambda\left[r\mu_{r-1} + \frac{d\mu_r}{d\lambda}\right]$$

para $r = 1, 2, 3, \ldots$. También, use esta fórmula recursiva y el hecho que $\mu_0 = 1$ y $\mu_1 = 0$ para encontrar μ_2, μ_3 y μ_4, y verifique la fórmula dada para α_3 en el ejercicio 5.52.

5.54 Use el teorema 5.9 para encontrar la función generatriz de momentos de $Y = X - \lambda$, donde X es una variable aleatoria que tiene la distribución de Poisson con el parámetro λ, y úsela para verificar que $\sigma_X^2 = \lambda$.

APLICACIONES

5.55 Si la probabilidad es 0.75 de que una persona creerá un rumor acerca de los delitos de cierto político, encuentre las probabilidades de que

(a) la octava persona que escucha el rumor será la quinta en creerlo;

(b) la décima quinta persona que escuche el rumor será la décima en creerlo.

5.56 Si las probabilidades de tener un hijo o una hija son ambas 0.50, encuentre las probabilidades de que

(a) el cuarto niño de una familia sea el primer hijo varón;

(b) el séptimo niño de una familia sea su segunda hija;

(c) el décimo niño de una familia sea su cuarto o quinto hijo varón.

5.57 Una experta tiradora certera falla en dar en el blanco cinco por ciento de las veces. Encuentre la probabilidad de que fallará en dar en el blanco por segunda vez en el décimo quinto tiro, use

(a) la fórmula para la distribución binomial negativa;

(b) el teorema 5.5 y la tabla I.

5.58 Al grabar un comercial de televisión, la probabilidad es 0.30 de que cierto actor dirá correctamente sus líneas en una toma cualquiera. ¿Cuál es la probabilidad de que dirá correctamente sus líneas por primera vez en la sexta toma?

5.59 En una "prueba de tortura" un apagador de luz se prende y se apaga hasta que falla. Si la probabilidad es 0.001 de que el apagador falle en cualquier momento en que se prenda o se apague, ¿cuál es la probabilidad de que el apagador no falle durante las primeras 800 veces que se prende o se apaga? Suponga que se satisfacen las condiciones requeridas por la distribución geométrica y use logaritmos.

5.60 Adapte la fórmula del teorema 5.5 de manera que se pueda usar para expresar probabilidades geométricas en términos de probabilidades binomiales, y use la fórmula de la tabla I para

(a) verificar el resultado del ejemplo 5.5;

(b) rehacer el ejercicio 5.58.

5.61 Un ingeniero de control de calidad inspecciona una muestra tomada al azar de dos calculadoras manuales de cada lote que llega de tamaño 18 y acepta el lote si ambas están en buenas condiciones de trabajo; de otra manera, se inspecciona todo el lote y el costo se carga al vendedor. ¿Cuáles son las probabilidades de que un lote así se acepte sin inspección adicional si contiene

(a) cuatro calculadoras que no están en buenas condiciones de trabajo;

(b) ocho calculadoras que no están en buenas condiciones de trabajo;

(c) 12 calculadoras que no están en buenas condiciones de trabajo?

5.62 De los 16 solicitantes para un trabajo, 10 tienen título universitario. Si se escogen tres de los solicitantes al azar para entrevistas, ¿cuáles son las probabilidades de que

(a) ninguno tenga un título universitario;

(b) uno tenga un título universitario;

(c) dos tengan un título universitario;

(d) los tres tengan un título universitario?

5.63 Encuentre la media y la varianza de la distribución hipergeométrica con $n = 3$, $N = 16$ y $k = 10$, use

(a) los resultados del ejercicio 5.62;

(b) las fórmulas del teorema 5.7.

5.64 ¿Cuál es la probabilidad de que una auditora de impuestos sólo descubra dos declaraciones de impuestos con deducciones ilegítimas si selecciona al azar cinco declaraciones entre 15, de las cuales nueve contienen deducciones ilegítimas?

5.65 Verifique en cada caso si se satisface la condición para la aproximación binomial a la distribución hipergeométrica:

(a) $N = 200$ y $n = 12$;

(b) $N = 500$ y $n = 20$;

(c) $N = 640$ y $n = 30$.

5.66 Un embarque de 80 alarmas contra robo contiene 4 que son defectuosas. Si del embarque se seleccionan al azar tres y se embarcan a un cliente, encuentre la probabilidad de que el cliente recibirá exactamente una unidad mala, use

(a) la fórmula de la distribución hipergeométrica;

(b) la distribución binomial como una aproximación.

5.67 Entre los 300 empleados de una compañía, 240 son sindicalizados, mientras que los otros no lo son. Si se seleccionan al azar seis de los empleados para prestar su servicios en un comité que administra el fondo de pensión, encuentre la probabilidad de que cuatro de los seis serán sindicalizados, use

(a) la fórmula de la distribución hipergeométrica;

(b) la distribución binomial como una aproximación.

5.68 Un panel de 300 personas escogidas para servir de jurados incluye a 30 que tienen menos de 25 años de edad. Puesto que el jurado de 12 personas escogido de este panel para juzgar una delito de narcóticos no incluye a nadie menor de 25 años de edad, el abogado defensor de los jóvenes acusados se queja de que este jurado realmente no es representativo. Ciertamente, él argumenta, si la selección fuera al azar, la probabilidad de tener uno de los 12 jurados de menos de 25 años de edad debe ser *muchas veces* la probabilidad de no tener ninguno de ellos menor a 25 años de edad. En realidad, ¿cuál es la razón de estas dos probabilidades?

5.69 Verifique en cada caso si los valores de n y θ satisfacen la regla práctica para una buena aproximación, una excelente aproximación, o ninguna cuando queremos usar la distribución de Poisson para aproximar las probabilidades binomiales

(a) $n = 125$ y $\theta = 0.10$;

(b) $n = 25$ y $\theta = 0.04$;

(c) $n = 120$ y $\theta = 0.05$;

(d) $n = 40$ y $\theta = 0.06$.

5.70 Con respecto al ejercicio 5.8, determine el valor de x (entre 5 y 15) para el cual el porcentaje de error es el más grande cuando usamos la distribución de Poisson con $\lambda = 7.5$ para aproximar la distribución binomial con $n = 150$ y $\theta = 0.05$.

5.71 Por experiencia se sabe que 1.4 por ciento de las llamadas recibidas en un conmutador son números equivocados. Use la aproximación de Poisson a la distribución binomial para determinar la probabilidad que de las 150 llamadas que recibe el conmutador dos son números equivocados.

5.72 Los registros muestran que la probabilidad es 0.0012 de que una persona se intoxicará con alimentos si pasa el día en cierta feria estatal. Use la aproximación de Poisson a la distribución binomial para encontrar la probabilidad de que entre 1,000 personas que asisten a la feria estatal cuando mucho dos se intoxicarán por alimentos.

5.73 En una ciudad dada, 4 por ciento de todos los conductores con licencia estarán involucrados en al menos un accidente automovilístico en un año dado cualquiera. Use la aproximación de Poisson a la distribución binomial para determinar la probabilidad de que entre 150 conductores con licencia escogidos al azar en esta ciudad

(a) sólo cinco estarán involucrados en al menos un accidente en un año dado cualquiera;

(b) cuando mucho tres estarán involucrados en al menos un accidente en un año dado cualquiera.

5.74 Con respecto al ejemplo 5.13 y la impresión de computadora de la figura 5.4, encuentre la probabilidad de que una lámina del metal de 15 pies cuadrados tendrá entre ocho y doce defectos, use

(a) los valores en la columna P(X = K);

(b) los valores en la columna P(X LESS OR = K).

5.75 El número de quejas que un negocio de tintorería recibe por día es una variable aleatoria que tiene la distribución de Poisson con $\lambda = 3.3$. Use la fórmula para la distribución de Poisson para encontrar la probabilidad de que sólo recibirá dos quejas en un día dado cualquiera.

5.76 El número de descomposturas mensuales de una computadora es una variable aleatoria que tiene una distribución de Poisson con $\lambda = 1.8$. Use la fórmula de la distribución de Poisson para encontrar las probabilidades de que esta computadora funcionará por un mes

(a) sin descomposturas;

(b) con sólo una descompostura.

5.77 Use la tabla II para verificar los resultados del ejercicio 5.76.

5.78 En cierta región desértica el número de personas que se enferman gravemente cada año por comer cierta planta venenosa es una variable aleatoria que tiene la distribución de Poisson con $\lambda = 5.2$. Use la tabla II para encontrar las probabilidades de

(a) tres enfermedades como ésa en un año dado;

(b) al menos 10 enfermedades como ésa en un año dado;

(c) cualquier número entre cuatro y seis enfermedades como ésa en un año dado.

5.79 En la inspección de una tela producida en rollos continuos, el número de imperfecciones por yarda es una variable aleatoria que tiene la distribución de Poisson con $\lambda = 0.25$. Encuentre la probabilidad de que 2 yardas de la tela tendrá cuando mucho una imperfección, use

(a) la tabla II;

(b) la impresión de computadora de la figura 5.5.

5.80 (a) Use un programa de cómputo para calcular la probabilidad *exacta* de obtener uno o más defectuosos en una muestra de tamaño 100 tomada de un lote de 1,000 productos fabricados que se supone contiene seis defectuosos.

(b) Aproxime esta probabilidad mediante el uso de la distribución binomial apropiada.

(c) Aproxime esta probabilidad mediante el uso de la distribución de Poisson apropiada y compare los resultados de los incisos (a), (b) y (c).

5.8 LA DISTRIBUCIÓN MULTINOMIAL

Una generalización inmediata de la distribución binomial surge cuando cada ensayo tiene más de dos resultados posibles, las probabilidades de los resultados correspondientes son las mismas para cada ensayo, y los ensayos son todos independientes. Este sería el caso, por ejemplo, cuando a las personas entrevistadas en una encuesta de opinión se les pregunta si están a favor de una candidato, contra él o indecisos, o cuando las muestras de productos manufacturados se clasifican como excelente, arriba del promedio, en el promedio o inferior.

Para tratar esta clase de problemas en general, consideremos el caso donde hay n ensayos independientes que permiten k resultados mutuamente excluyentes cuyas probabilidades respectivas son $\theta_1, \theta_2, \ldots, \theta_k \left(\text{con } \sum_{i=1}^{k} \theta_i = 1\right)$. Al referirnos a los resultados como que son de la primera clase, la segunda clase,... y la $k$ésima clase, estaremos interesados en la probabilidad de obtener x_1 resultados de la primera clase, x_2 resultados de la segunda clase,... y x_k resultados de la $k$ésima clase $\left(\text{con } \sum_{i=1}^{k} x_i = n\right)$.

Al proceder como en la derivación de la fórmula para la distribución binomial, primero encontramos la probabilidad de obtener x_1 resultados de la primera clase,

x_2 resultados de la segunda clase, ... y x_k resultados de la $k$ésima clase *en un orden específico* es $\theta_1^{x_1} \cdot \theta_2^{x_2} \cdot \ldots \cdot \theta_k^{x_k}$. Para tener la probabilidad correspondiente para todos esos resultados de cada clase *en cualquier orden*, tendremos que multiplicar la probabilidad para cualquier orden específico por

$$\binom{n}{x_1, x_2, \ldots, x_k} = \frac{n!}{x_1! \cdot x_2! \cdot \ldots \cdot x_k!}$$

de acuerdo al teorema 1.8.

DEFINICIÓN 5.8 Las variables aleatorias $X_1, X_2, \ldots, X_n$ tienen una **distribución multinomial** y se conocen como variables aleatorias multinomiales si y sólo si su distribución de probabilidad conjunta está dada por

$$f(x_1, x_2, \ldots, x_k; n, \theta_1, \theta_2, \ldots, \theta_k) = \binom{n}{x_1, x_2, \ldots, x_k} \cdot \theta_1^{x_1} \cdot \theta_2^{x_2} \cdot \ldots \cdot \theta_k^{x_k}$$

para $x_i = 0, 1, \ldots, n$ para cada i, donde $\sum_{i=1}^{k} x_i = n$ y $\sum_{i=1}^{k} \theta_i = 1$.

Así, los números de resultados de las diferentes clases son variables aleatorias que tienen la distribución multinomial con los parámetros $n, \theta_1, \theta_2, \ldots$ y θ_k. El nombre "multinomial" se deriva del hecho de que para los valores de x_i, las probabilidades son iguales a los términos correspondientes de la expansión multinomial de $(\theta_1 + \theta_2 + \cdots + \theta_k)^n$.

EJEMPLO 5.14

Cierta ciudad tiene tres estaciones de televisión. Durante el horario principal del sábado en la noche, el canal 12 tiene 50 por ciento del público, el canal 10 tiene 30 por ciento del público y el canal 3 tiene 20 por ciento. Encuentre la probabilidad de que entre ocho personas que ven televisión en esa ciudad, escogidas al azar un sábado por la noche, cinco estarán viendo el canal 12, dos estarán viendo el canal 10 y una estará viendo el canal 3.

Solución

Sustituimos $x_1 = 5, x_2 = 2, x_3 = 1, \theta_1 = 0.50, \theta_2 = 0.30, \theta_3 = 0.20$ y $n = 8$ en la fórmula de la definición 5.8, y obtenemos

$$f(5, 2, 1; 8, 0.50, 0.30, 0.20) = \frac{8!}{5! \cdot 2! \cdot 1!} (0.50)^5 (0.30)^2 (0.20)$$

$$= 0.0945 \qquad \blacktriangle$$

5.9 LA DISTRIBUCIÓN HIPERGEOMÉTRICA MULTIVARIADA

Justo como la distribución hipergeométrica toma el lugar de la distribución binomial para el muestreo sin reemplazo, también existe una distribución multivariada análoga a la distribución multinomial que aplica al muestreo sin reemplazo. Para derivar esta fórmula, consideremos un conjunto de N elementos, de los cuales M_1 son elementos de la primera clase, M_2 son elementos de la segunda clase, ... y M_k son elementos de la $k$ésima clase tales que $\sum_{i=1}^{k} M_i = N$. Como en relación con la distribución multinomial, estamos interesados en la probabilidad de obtener x_1 elementos (resultados) de la primera clase, x_2 elementos de la segunda clase,..., y x_k elementos de la $k$ésima clase, pero ahora estamos escogiendo sin reemplazo, n de los N elementos del conjunto.

Hay $\binom{M_1}{x_1}$ maneras de escoger x_1 de los M_1 elementos de la primera clase, $\binom{M_2}{x_2}$ maneras de escoger x_2 elementos de los M_2 elementos de la segunda clase, ... y $\binom{M_k}{x_k}$ maneras de escoger x_k elementos de los M_k elementos de la $k$ésima clase, y por tanto, $\binom{M_1}{x_1}\binom{M_2}{x_2}\cdot \ldots \cdot\binom{M_k}{x_k}$ maneras de escoger los $\sum_{i=1}^{k} x_i = n$ elementos requeridos. Puesto que hay $\binom{N}{n}$ maneras de escoger n de los N elementos en el conjunto y suponemos que todas son igualmente posibles (que es lo que queremos decir cuando afirmamos que la selección es al azar), se sigue que la probabilidad deseada está dada por $\binom{M_1}{x_1}\binom{M_2}{x_2}\cdot \ldots \cdot\binom{M_k}{x_k}\Big/\binom{N}{n}$.

DEFINICIÓN 5.9 Las variables aleatorias $X_1, X_2, \ldots, X_k$ tienen una **distribución hipergeométrica multivariada** y se conocen como variables aleatorias hipergeométricas multivariadas si y sólo si su distribución de probabilidad conjunta está dada por

$$f(x_1, x_2, \ldots, x_k; n, M_1, M_2, \ldots, M_k) = \frac{\binom{M_1}{x_1}\binom{M_2}{x_2}\cdot \ldots \cdot\binom{M_k}{x_k}}{\binom{N}{n}}$$

para $x_i = 0, 1, \ldots, n$ y $x_i \leqq M_i$ para cada i, donde $\sum_{i=1}^{k} x_i = n$ y $\sum_{i=1}^{k} M_i = N$.

Así, la distribución conjunta de las variables aleatorias bajo consideración, esto es, la distribución de los números de resultados de las diferentes clases, es una distribución hipergeométrica multivariada con los parámetros $n, M_1, M_2, \ldots$ y M_k.

EJEMPLO 5.15

En un panel de presuntos jurados participan seis hombres casados, tres hombres solteros, siete mujeres casadas y cuatro mujeres solteras. Si la selección es al azar, ¿cuál es la probabilidad de que el jurado consistirá de cuatro hombres casados, un hombre soltero, cinco mujeres casadas y dos mujeres solteras?

Solución

Sustituimos $x_1 = 4, x_2 = 1, x_3 = 5, x_4 = 2, M_1 = 6, M_2 = 3, M_3 = 7, M_4 = 4$, $N = 20$ y $n = 12$ en la fórmula de la definición 5.9, y obtenemos

$$f(4, 1, 5, 2; 12, 6, 3, 7, 4) = \frac{\binom{6}{4}\binom{3}{1}\binom{7}{5}\binom{4}{2}}{\binom{20}{12}}$$

$$= 0.0450 \qquad \blacktriangle$$

EJERCICIOS

5.81 Si $X_1, X_2, \ldots, X_k$ tienen la distribución multinomial de la definición 5.8, demuestre que la media de la distribución marginal de X_i es $n\theta_i$ para $i = 1, 2, \ldots, k$.

5.82 Si $X_1, X_2, \ldots, X_k$ tienen la distribución multinomial de la definición 5.8, demuestre que la covarianza de X_i y X_j es $-n\theta_i\theta_j$ para $i = 1, 2, \ldots, k$, $j = 1, 2, \ldots, k$ e $i \neq j$.

APLICACIONES

5.83 Las probabilidades son 0.40, 0.50 y 0.10 de que, en el tránsito de la ciudad, cierta clase de auto compacto promediará menos de 22 millas por galón, de 22 a 26 millas por galón o más de 26 millas por galón. Encuentre la probabilidad de que de 10 autos probados de éstos tres promediarán menos de 22 millas por galón, seis promediarán de 22 a 26 millas por galón y uno promediará 26 millas por galón.

5.84 Suponga que las probabilidades son 0.60, 0.20, 0.10 y 0.10 de que una declaración de impuestos estatales a la renta se llenará correctamente, que sólo contendrá errores en favor del causante, que sólo contendrá errores en favor del estado o que contendrá ambas clases de errores. ¿Cuál es la probabilidad de que de 12 de tales declaraciones de impuesto a la renta escogidas al azar para auditoría, cinco estarán correctamente llenadas, cuatro sólo contendrán errores a favor del causante, dos sólo contendrán errores a favor del estado y uno contendrá ambas clases de errores?

5.85 De acuerdo a la teoría mendeliana de la herencia, si plantas con semillas amarillas redondas se cruzan con plantas con semillas verdes arrugadas, las probabilidades de obtener una planta que produzca semillas amarillas redondas, semillas amarillas arrugadas, semillas verdes redondas o semillas verdes arrugadas son, respectivamente, $\frac{9}{16}$, $\frac{3}{16}$, $\frac{3}{16}$ y $\frac{1}{16}$. ¿Cuál es la probabilidad de que de nueve de esas plantas obtenidas habrá cuatro que produzcan semillas amarillas redondas, dos que produzcan semillas amarillas arrugadas, tres que produzcan semillas verdes redondas y ninguna que produzca semillas verdes arrugadas?

5.86 Si entre 18 ladrillos de vidrio defectuosos hay 10 que tienen fracturas pero no decoloraciones, cinco que tienen decoloraciones pero no fracturas y tres que tienen fracturas y decoloraciones, ¿cuál es la probabilidad de que de seis de los ladrillos (escogidos al azar para pruebas adicionales) tres tendrán fracturas pero no decoloraciones, uno tendrá decoloraciones pero no fracturas y dos tendrán fracturas y decoloraciones?

5.87 De 25 dólares de plata acuñados en 1903 hay 15 de la casa de moneda de Filadelfia, siete de la casa de moneda de Nueva Orleans y tres de la casa de moneda de San Francisco. Si se escogen al azar cinco de estos dólares de plata, encuentre las probabilidades de obtener

(a) cuatro de la casa de moneda de Filadelfia y uno de la casa de moneda de Nueva Orleans;

(b) tres de la casa de moneda de Filadelfia y uno de cada una de las otras dos casas de moneda.

REFERENCIAS

Se puede encontrar información útil sobre las diversas distribuciones de probabilidad especiales en

DERMAN, C., GLESER, L., and OLKIN, I., *Probability Models and Applications*. Nueva York: Macmillan Publishing Co., Inc., 1980,

HASTINGS, N. A. J., and PEACOCK, J. B., *Statistical Distributions*. Londres: Butterworth & Co. Ltd., 1975,

y

JOHNSON, N. L., and KOTZ, S., *Discrete Distributions*. Boston: Houghton Mifflin Company, 1969.

Se pueden encontrar las probabilidades binominales para $n = 2$ a $n = 49$ en

Tables of the Binomial Probability Distribution, National Bureau of Standards Applied Mathematics Series No. 6, Washington, D.C.: U.S. Government Printing Office, 1950,

y para $n = 50$ a $n = 100$ en

ROMIG, H. G., *50–100 Binomial Tables*. Nueva York: John Wiley & Sons, Inc., 1953.

La tabla de probabilidades de Poisson que más ampliamente se usa es

MOLINA, E. C., *Poisson's Exponential Binomial Limit*. Melbourne, Fla.: Robert E. Krieger Publishing Company, 1973 Reprint.

CAPÍTULO

6

Densidades de probabilidad especiales

6.1 INTRODUCCIÓN

En este capítulo estudiaremos algunas de las densidades de probabilidad que figuran de manera muy prominente en la teoría estadística y en las aplicaciones. Además de las tratadas en el texto, se introducen otras más en los ejercicios posteriores a la sección 6.4, y tres de las densidades de probabilidad que son de importancia básica en la teoría de muestreo se abordarán en el capítulo 8. Como en el capítulo 5, derivaremos parámetros y funciones generatrices de momentos, dejando, una vez más, algunos detalles como ejercicios.

6.2 LA DISTRIBUCIÓN UNIFORME

Las densidades de probabilidad de los ejemplos 3.8 y 3.11 son casos especiales de la **distribución uniforme**; en la figura 3.7 se muestra la gráfica de la del ejemplo 3.11. En general,

DEFINICIÓN **6.1** Una variable aleatoria tiene una **distribución uniforme** y se conoce como una variable aleatoria uniforme continua si y sólo si su densidad de probabilidad está dada por

$$u(x; \alpha, \beta) = \begin{cases} \dfrac{1}{\beta - \alpha} & \text{para } \alpha < x < \beta \\ 0 & \text{en cualquier otra parte} \end{cases}$$

Los parámetros α y β de esta densidad de probabilidad son constantes reales, con $\alpha < \beta$, y se puede visualizar en la figura 6.1. En el ejercicio 6.2 se pedirá al lector que verifique que

TEOREMA 6.1 La media y la varianza de la distribución uniforme están dadas por

$$\mu = \frac{\alpha + \beta}{2} \quad \text{y} \quad \sigma^2 = \frac{1}{12}(\beta - \alpha)^2$$

Aunque la distribución uniforme tiene algunas aplicaciones directas, una de las cuales se examinará en el ejemplo 7.8, su valor principal es que, a causa de su simplicidad, se presta rápidamente a la tarea de ilustrar diversos aspectos de la teoría estadística.

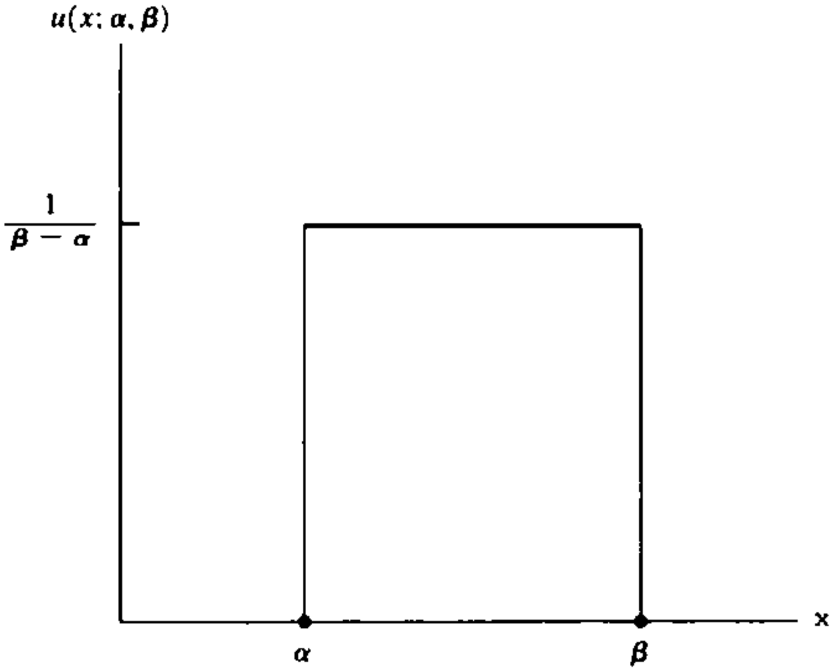

Figura 6.1 La distribución uniforme.

6.3 LAS DISTRIBUCIONES GAMMA, EXPONENCIAL Y JI CUADRADA

Algunos de los ejemplos y ejercicios de los capítulos 3 y 4 atañen a variables aleatorias que tienen densidades de probabilidad de la forma

$$f(x) = \begin{cases} kx^{\alpha-1}e^{-x/\beta} & \text{para } x > 0 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

donde $\alpha > 0$, $\beta > 0$ y k debe ser tal que el área total bajo la curva sea igual a 1. Para evaluar k, primero hacemos la substitución $y = \frac{x}{\beta}$, lo cual nos da

$$\int_0^\infty kx^{\alpha-1}e^{-x/\beta}\,dx = k\beta^\alpha \int_0^\infty y^{\alpha-1}e^{-y}\,dy$$

La integral así obtenida depende sólo de α y define la bien conocida **función** gamma

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1}e^{-y}\,dy \qquad \text{para } \alpha > 0$$

que se trata en detalle en muchos de los textos de cálculo avanzado. Al integrar por partes, lo que se deja al lector en el ejercicio 6.7, encontramos que la función gamma satisface la fórmula recursiva

$$\Gamma(\alpha) = (\alpha - 1)\cdot\Gamma(\alpha - 1)$$

para $\alpha > 1$, y puesto que

$$\Gamma(1) = \int_0^\infty e^{-y}\,dy = 1$$

se sigue por la aplicación repetida de la fórmula recursiva que $\Gamma(\alpha) = (\alpha - 1)!$ donde α es un entero positivo. También, un valor especial importante es $\Gamma(\frac{1}{2}) = \sqrt{\pi}$, como se le pedirá al lector que verifique en el ejercicio 6.9.

Regresamos ahora al problema de evaluar k, igualamos la integral obtenida a 1, y obtenemos

$$\int_0^\infty kx^{\alpha-1}e^{-x/\beta}\,dx = k\beta^\alpha\Gamma(\alpha) = 1$$

y por tanto

$$k = \frac{1}{\beta^\alpha\Gamma(\alpha)}$$

Esto nos lleva a la siguiente definición de la distribución gamma.

DEFINICIÓN 6.2 Una variable aleatoria X tiene una **distribución gamma** y se conoce como una variable aleatoria gamma si y sólo si su densidad de probabilidad está dada por

$$g(x;\alpha,\beta) = \begin{cases} \dfrac{1}{\beta^\alpha\Gamma(\alpha)}x^{\alpha-1}e^{-x/\beta} & \text{para } x > 0 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

donde $\alpha > 0$ y $\beta > 0$.

Cuando α no es un entero positivo, el valor de $\Gamma(\alpha)$ deberá buscarse en una tabla especial. Para dar al lector una idea sobre la forma de las gráficas de las densidades gamma, en la figura 6.2 se muestran algunas para varios valores especiales de α y β.

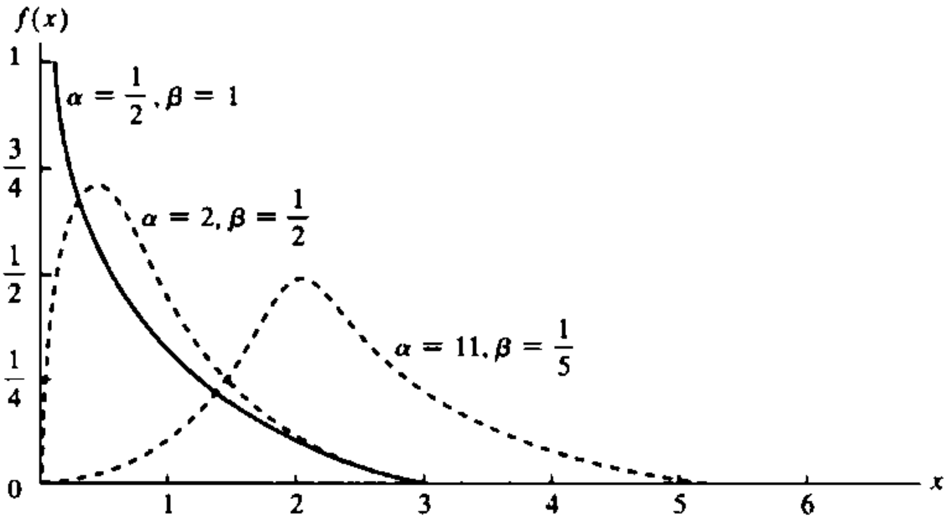

Figura 6.2 Gráficas de distribuciones gamma.

Algunos casos especiales de la distribución gamma juegan papeles importantes en la estadística, para $\alpha = 1$ y $\beta = \theta$, obtenemos

DEFINICIÓN 6.3 Una variable aleatoria X tiene una **distribución exponencial** y se conoce como una variable aleatoria exponencial si y sólo si su densidad de probabilidad está dada por

$$g(x; \theta) = \begin{cases} \frac{1}{\theta} e^{-x/\theta} & \text{para } x > 0 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

donde $\theta > 0$.

La densidad se visualiza en la figura 6.3

Para demostrar cómo puede surgir en la práctica una distribución exponencial, vayamos a la situación descrita en el ejercicio 5.48, donde nos interesa la probabilidad de obtener x éxitos durante un intervalo de tiempo de longitud t cuando (i) la probabilidad de un éxito durante un intervalo de tiempo muy pequeño de t a $t + \Delta t$ es $\alpha \cdot \Delta t$, (ii) la probabilidad de más de un éxito durante un intervalo de tiempo tal es insignificante, y (iii) la probabilidad de un éxito durante un intervalo de tiempo tal no depende de lo que haya pasado antes del tiempo t. En ese ejercicio, demostramos que el número de éxitos es un valor de la variable aleatoria discreta X que tiene la distribución de Poisson con $\lambda = \alpha t$. Ahora determinemos la densidad de probabilidad de la variable aleatoria continua Y, el **tiempo de espera** hasta el primer éxito. Claramente:

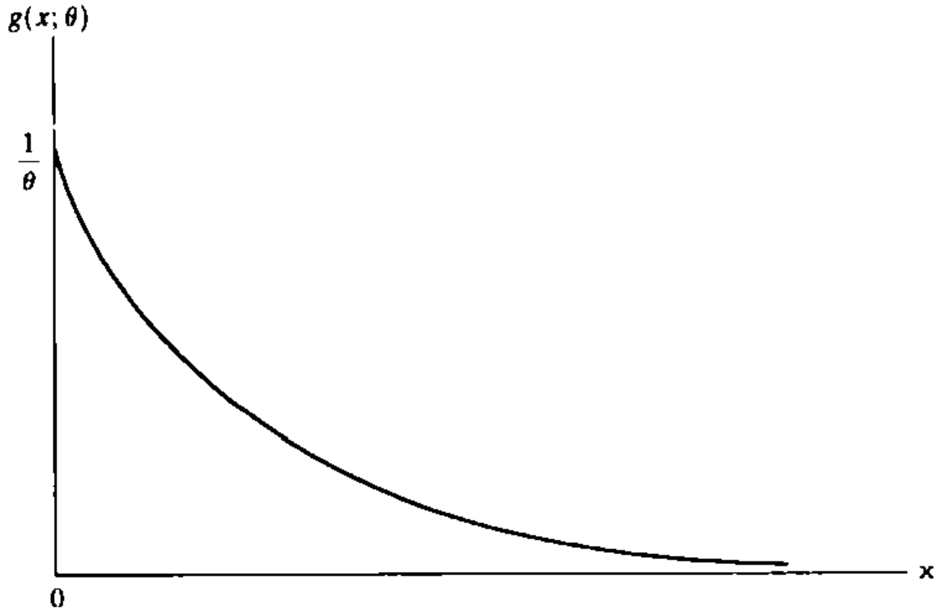

Figura 6.3 Distribución exponencial.

$$
\begin{aligned}
F(y) = P(Y \leqq y) &= 1 - P(Y > y) \\
&= 1 - P(0 \text{ éxitos en el intervalo de tiempo de longitud } y) \\
&= 1 - p(0; \alpha y) \\
&= 1 - \frac{e^{-\alpha y}(\alpha y)^0}{0!} \\
&= 1 - e^{-\alpha y} \qquad \text{para } y > 0
\end{aligned}
$$

y $F(y) = 0$ para $y \leqq 0$. Una vez que hemos encontrado así la función de distribución de Y, encontramos que la diferenciación con respecto a y nos da

$$
f(y) = \begin{cases} \alpha e^{-\alpha y} & \text{para } y > 0 \\ 0 & \text{en cualquier otra parte} \end{cases}
$$

que es la distribución exponencial con $\theta = \dfrac{1}{\alpha}$.

La distribución exponencial se aplica no sólo a la ocurrencia del primer éxito en un **proceso de Poisson**, que es como llamamos a una situación como la descrita en el ejercicio 5.48, pero en virtud de la condición (iii) (véase el ejercicio 6.16), se aplica también a los tiempos de espera entre éxitos.

EJEMPLO 6.1

En una cierta localidad en la carretera I-10, el número de autos que exceden el límite de velocidad en más de 10 millas por hora en media hora es una variable aleatoria que tiene una distribución de Poisson con $\lambda = 8.4$. ¿Cuál es la probabilidad de un tiempo de espera menor de 5 minutos entre autos que exceden el límite de velocidad en más de 10 millas por hora?

Solución

Al usar media hora como la unidad de tiempo, tenemos $\alpha = \lambda = 8.4$. Por consiguiente, el tiempo de espera es una variable aleatoria que tiene una distribución exponencial con $\theta = \frac{1}{8.4}$ y, puesto que 5 minutos es $\frac{1}{6}$ de la unidad de tiempo, encontramos que la probabilidad deseada es

$$\int_0^{1/6} 8.4e^{-8.4x}\,dx = -e^{-8.4x}\Big|_0^{1/6} = -e^{-1.4} + 1$$

que es aproximadamente 0.75 ▲

Otro caso especial de la distribución gamma surge cuando $\alpha = \frac{\nu}{2}$ y $\beta = 2$, donde ν es la minúscula de la letra griega nu.

DEFINICIÓN 6.4 Una variable aleatoria X tiene una **distribución ji cuadrada** y se conoce como una variable aleatoria ji cuadrada si y sólo si su densidad de probabilidad está dada por

$$f(x) = \begin{cases} \dfrac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{\frac{\nu-2}{2}} e^{-\frac{x}{2}} & \text{para } x > 0 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

El parámetro ν se conoce como el **número de grados de libertad**, o simplemente **grados de libertad**. La distribución ji cuadrada juega un papel muy importante en la teoría del muestreo, y se examina con cierto detalle en el capítulo 8.

Para derivar las fórmulas para la media y la varianza de la distribución gamma y, por tanto, de las distribuciones exponencial y ji cuadrada, demostremos primero el siguiente teorema.

TEOREMA 6.2 El $r$ésimo momento alrededor del origen de la distribución gamma está dado por

$$\mu_r' = \frac{\beta^r \Gamma(\alpha + r)}{\Gamma(\alpha)}$$

Demostración. Por la definición 4.2,

$$\mu_r' = \int_0^{\infty} x^r \cdot \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}\,dx = \frac{\beta^r}{\Gamma(\alpha)} \cdot \int_0^{\infty} y^{\alpha+r-1} e^{-y}\,dy$$

donde hacemos $y = \frac{x}{\beta}$. Puesto que la integral de la derecha es $\Gamma(r + \alpha)$ de acuerdo a la definición de la función gamma en la página 205, esto completa la demostración. ▼

Con el uso de este teorema, derivemos ahora los siguientes resultados sobre la distribución gamma.

TEOREMA 6.3 La media y la varianza de la distribución gamma están dadas por

$$\mu = \alpha\beta \quad \text{y} \quad \sigma^2 = \alpha\beta^2$$

Demostración. Por el teorema 6.2 con $r = 1$ y $r = 2$, obtenemos

$$\mu_1' = \frac{\beta\Gamma(\alpha + 1)}{\Gamma(\alpha)} = \alpha\beta$$

y

$$\mu_2' = \frac{\beta^2\Gamma(\alpha + 2)}{\Gamma(\alpha)} = \alpha(\alpha + 1)\beta^2$$

así que $\mu = \alpha\beta$ y $\sigma^2 = \alpha(\alpha + 1)\beta^2 - (\alpha\beta)^2 = \alpha\beta^2$. ▼

Al sustituir en estas fórmulas $\alpha = 1$ y $\beta = \theta$ para la distribución exponencial y $\alpha = \frac{\nu}{2}$ y $\beta = 2$ para la distribución ji cuadrada, obtenemos

COROLARIO 1 La media y la varianza de la distribución exponencial están dadas por

$$\mu = \theta \quad \text{y} \quad \sigma^2 = \theta^2$$

COROLARIO 2 La media y la varianza de la distribución ji cuadrada están dadas por

$$\mu = \nu \quad \text{y} \quad \sigma^2 = 2\nu$$

Para referencia futura, también demos aquí la función generatriz de momentos de la distribución gamma

TEOREMA 6.4 La función generatriz de momentos de la distribución gamma está dada por

$$M_X(t) = (1 - \beta t)^{-\alpha}$$

Se pedirá al lector que demuestre este resultado y lo use para encontrar algunos de los momentos más bajos en los ejercicios 6.12 y 6.13.

6.4 LA DISTRIBUCIÓN BETA

La densidad uniforme $f(x) = 1$ para $0 < x < 1$ y $f(x) = 0$ en cualquier otra parte es un caso especial de la **distribución beta**, la cual se define de la siguiente manera

DEFINICIÓN 6.5 Una variable aleatoria X tiene una **distribución beta** y se conoce como una variable aleatoria beta si y sólo si su densidad de probabilidad está dada por

$$f(x) = \begin{cases} \dfrac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \cdot \Gamma(\beta)} x^{\alpha-1}(1 - x)^{\beta-1} & \text{para } 0 < x < 1 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

donde $\alpha > 0$ y $\beta > 0$.

En años recientes, la distribución beta ha encontrado aplicaciones importantes en la **inferencia bayesiana**, donde los parámetros se consideran como variables aleatorias, y hay necesidad de una densidad de probabilidad bastante "flexible" para el parámetro θ de la distribución binomial, el cual sólo toma valores distintos a cero en el intervalo desde 0 hasta 1. Con "flexible" queremos decir que la densidad de probabilidad puede tomar una gran variedad de formas diferentes, como se pedirá al lector que verifique para la distribución beta en el ejercicio 6.27. Este uso de la distribución beta se examina en el capítulo 10.

No demostraremos aquí que el área total bajo la curva de la distribución beta, como la de cualquier densidad de probabilidad, es igual a 1, pero en la demostración del teorema que sigue, nos valdremos del hecho que

$$\int_0^1 \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \cdot \Gamma(\beta)} x^{\alpha-1}(1 - x)^{\beta-1}\, dx = 1$$

y por tanto que

$$\int_0^1 x^{\alpha-1}(1 - x)^{\beta-1}\, dx = \frac{\Gamma(\alpha) \cdot \Gamma(\beta)}{\Gamma(\alpha + \beta)}$$

Esta integral define la **función beta**, cuyos valores se denotan $B(\alpha, \beta)$; en otras palabras, $B(\alpha, \beta) = \dfrac{\Gamma(\alpha) \cdot \Gamma(\beta)}{\Gamma(\alpha + \beta)}$. En cualquier libro de texto de cálculo avanzado se puede encontrar un análisis detallado de la función beta.

TEOREMA 6.5 La media y la varianza de la distribución beta están dadas por

$$\mu = \frac{\alpha}{\alpha + \beta} \quad \text{y} \quad \sigma^2 = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

Demostración. Por definición,

$$\begin{aligned}
\mu &= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \cdot \Gamma(\beta)} \cdot \int_0^1 x \cdot x^{\alpha - 1}(1 - x)^{\beta - 1}\, dx \\
&= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \cdot \Gamma(\beta)} \cdot \frac{\Gamma(\alpha + 1) \cdot \Gamma(\beta)}{\Gamma(\alpha + \beta + 1)} \\
&= \frac{\alpha}{\alpha + \beta}
\end{aligned}$$

donde reconocimos la integral como $B(\alpha + 1, \beta)$ y usamos el hecho que $\Gamma(\alpha + 1) = \alpha \cdot \Gamma(\alpha)$ y $\Gamma(\alpha + \beta + 1) = (\alpha + \beta) \cdot \Gamma(\alpha + \beta)$. Pasos similares, los que se dejarán al lector en el ejercicio 6.28, dan

$$\mu_2' = \frac{(\alpha + 1)\alpha}{(\alpha + \beta + 1)(\alpha + \beta)}$$

y se sigue que

$$\begin{aligned}
\sigma^2 &= \frac{(\alpha + 1)\alpha}{(\alpha + \beta + 1)(\alpha + \beta)} - \left(\frac{\alpha}{\alpha + \beta}\right)^2 \\
&= \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} \qquad \blacktriangledown
\end{aligned}$$

EJERCICIOS

6.1 Demuestre que si una variable aleatoria tiene una densidad uniforme con los parámetros α y β, la probabilidad que asumirá un valor menor que $\alpha + p(\beta - \alpha)$ es igual a p.

6.2 Demuestre el teorema 6.1

6.3 Si una variable aleatoria X tiene una densidad uniforme con los parámetros α y β, encuentre su función de distribución.

6.4 Muestre que si una variable aleatoria tiene una densidad uniforme con los parámetros α y β, el $r$ésimo momento alrededor de la media es igual a

(a) 0 cuando r es impar;

(b) $\dfrac{1}{r + 1}\left(\dfrac{\beta - \alpha}{2}\right)^r$ cuando r es par.

6.5 Use los resultados del ejercicio 6.4 para encontrar α_3 y α_4 para la densidad uniforme con los parámetros α y β.

6.6 Se dice que una variable aleatoria tiene la **distribución de Cauchy** si su densidad está dada por

$$f(x) = \frac{\frac{\beta}{\pi}}{(x-\alpha)^2 + \beta^2} \qquad \text{para } -\infty < x < \infty$$

Demuestre que para esta distribución no existen μ_1' y μ_2'.

6.7 Use la integración por partes para mostrar que $\Gamma(\alpha) = (\alpha - 1)\cdot\Gamma(\alpha - 1)$ para $\alpha > 1$.

6.8 Efectúe un cambio apropiado de variable para mostrar que la integral que define la función gamma se puede escribir como

$$\Gamma(\alpha) = 2^{1-\alpha}\cdot\int_0^\infty z^{2\alpha-1}e^{-\frac{1}{2}z^2}\,dz \qquad \text{para } \alpha > 0$$

6.9 Al usar la forma de la función gamma del ejercicio 6.8, podemos escribir

$$\Gamma(\tfrac{1}{2}) = \sqrt{2}\int_0^\infty e^{-\frac{1}{2}z^2}\,dz$$

y por tanto

$$[\Gamma(\tfrac{1}{2})]^2 = 2\left\{\int_0^\infty e^{-\frac{1}{2}x^2}\,dx\right\}\left\{\int_0^\infty e^{-\frac{1}{2}y^2}\,dy\right\} = 2\int_0^\infty\int_0^\infty e^{-\frac{1}{2}(x^2+y^2)}\,dx\,dy$$

Cambie a coordenadas polares para evaluar esta integral doble, y así demostrar que $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

6.10 Encuentre las probabilidades de que el valor de una variable aleatoria excederá a 4 si tiene una distribución gama con

(a) $\alpha = 2$ y $\beta = 3$;

(b) $\alpha = 3$ y $\beta = 4$.

6.11 Muestre que una distribución gamma con $\alpha > 1$ tiene un máximo relativo en $x = \beta(\alpha - 1)$. ¿Qué pasa cuando $0 < \alpha < 1$ y $\alpha = 1$?

6.12 Demuestre el teorema 6.4, haga la sustitución $y = x\left(\frac{1}{\beta} - t\right)$ en la integral que define $M_X(t)$.

6.13 Expanda la función generatriz de momentos de la distribución gamma como una serie binomial, y lea los valores de μ_1', μ_2', μ_3' y μ_4'.

6.14 Use los resultados del ejercicio 6.13 para encontrar α_3 y α_4 para la distribución gamma.

6.15 Muestre que si una variable aleatoria tiene una densidad exponencial con el parámetro θ, la probabilidad de que asumirá un valor menor que $-\theta\cdot\ln(1 - p)$ es igual a p para $0 \leq p < 1$.

6.16 Si X tiene una distribución exponencial, demuestre que

$$P(X \geq t + T \mid X \geq T) = P(X \geq t)$$

Esta propiedad de una variable aleatoria exponencial es paralela a la de una variable aleatoria geométrica dada en el ejercicio 5.40.

6.17 Si X es una variable aleatoria que tiene una distribución exponencial con el parámetro θ, use los teoremas 4.10 y 6.4 para encontrar la función generatriz de momentos de la variable aleatoria $Y = X - \theta$.

6.18 Con respecto al ejercicio 6.17, use el hecho que los momentos de Y alrededor del origen son los momentos correspondientes de X alrededor de la media, encuentre α_3 y α_4 para la distribución exponencial con el parámetro θ.

6.19 Muestre que si $\nu > 2$, la distribución ji cuadrada tiene un máximo relativo en $x = \nu - 2$. ¿Qué sucede cuando $\nu = 2$ o $0 < \nu < 2$?

6.20 Una variable aleatoria X tiene la **distribución de Rayleigh** si y sólo si su densidad de probabilidad está dada por

$$f(x) = \begin{cases} 2\alpha x e^{-\alpha x^2} & \text{para } x > 0 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

donde $\alpha > 0$. Demuestre que para esta distribución

(a) $\mu = \dfrac{1}{2}\sqrt{\dfrac{\pi}{\alpha}}$;

(b) $\sigma^2 = \dfrac{1}{\alpha}\left(1 - \dfrac{\pi}{4}\right)$.

6.21 Una variable aleatoria X tiene una **distribución de Pareto** si y sólo si su densidad de probabilidad está dada por

$$f(x) = \begin{cases} \dfrac{\alpha}{x^{\alpha+1}} & \text{para } x > 1 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

donde $\alpha > 0$. Demuestre que μ_r' existe sólo si $r < \alpha$.

6.22 Con respecto al ejercicio 6.21, demuestre que para la distribución de Pareto $\mu = \dfrac{\alpha}{\alpha - 1}$ siempre que $\alpha > 1$.

6.23 Una variable aleatoria X tiene la **distribución de Weibull** si y sólo si su densidad de probabilidad está dada por

$$f(x) = \begin{cases} kx^{\beta-1}e^{-\alpha x^\beta} & \text{para } x > 0 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

donde $\alpha > 0$ y $\beta > 0$.

(a) Exprese k en términos de α y β.

(b) Demuestre que $\mu = \alpha^{-1/\beta}\Gamma\left(1 + \dfrac{1}{\beta}\right)$.

Observe que las distribuciones de Weibull con $\beta = 1$ son distribuciones exponenciales.

6.24 Si la variable aleatoria T es el tiempo para descompostura o defecto de un producto comercial y los valores de su densidad de probabilidad y su función de distribución en el tiempo t son $f(t)$ y $F(t)$, entonces su tasa de defecto en el tiempo t (véase también el ejercicio 5.41) está dada por $\frac{f(t)}{1 - F(t)}$. Así, la tasa de defecto en el tiempo t es la densidad de probabilidad de defecto en el tiempo t dado que no ocurre un defecto antes del tiempo t.

(a) Muestre que si T tiene una distribución exponencial, la tasa de defecto es constante.

(b) Muestre que si T tiene una distribución de Weibull (véase el ejercicio 6.23), la tasa de defecto está dada por $\alpha\beta t^{\beta-1}$.

6.25 Verifique que la integral de la densidad beta desde $-\infty$ hasta ∞ es igual a 1 para

(a) $\alpha = 2$ y $\beta = 4$;

(b) $\alpha = 3$ y $\beta = 3$.

6.26 Muestre que si $\alpha > 1$ y $\beta > 1$, la densidad beta tiene un máximo relativo en $x = \frac{\alpha - 1}{\alpha + \beta - 2}$.

6.27 Bosqueje las gráficas de las densidades beta que tienen

(a) $\alpha = 2$ y $\beta = 2$;

(b) $\alpha = \frac{1}{2}$ y $\beta = 1$;

(c) $\alpha = 2$ y $\beta = \frac{1}{2}$;

(d) $\alpha = 2$ y $\beta = 5$.

[*Sugerencia*: para evaluar $\Gamma(\frac{3}{2})$ y $\Gamma(\frac{5}{2})$, utilice la fórmula recursiva $\Gamma(\alpha) = (\alpha - 1) \cdot \Gamma(\alpha - 1)$ y el resultado del ejercicio 6.9.]

6.28 Verifique la expresión dada para μ_2' en la demostración del teorema 6.5.

6.29 Demuestre que los parámetros de la distribución beta se pueden expresar como sigue en términos de la media y la varianza de esta distribución:

(a) $\alpha = \mu\left[\frac{\mu(1 - \mu)}{\sigma^2} - 1\right]$;

(b) $\beta = (1 - \mu)\left[\frac{\mu(1 - \mu)}{\sigma^2} - 1\right]$.

6.30 Karl Pearson, uno de los fundadores de la estadística moderna, demostró que la ecuación diferencial

$$\frac{1}{f(x)} \cdot \frac{d[f(x)]}{dx} = \frac{d - x}{a + bx + cx^2}$$

nos da (para valores apropiados de las constantes a, b, c y d) la mayor parte de las distribuciones importantes de la estadística. Verifique que la ecuación diferencial da

(a) la distribución gamma cuando $a = c = 0$, $b > 0$ y $d > -b$;

(b) la distribución exponencial cuando $a = c = d = 0$ y $b > 0$;

(c) la distribución beta cuando $a = 0$, $b = -c$, $\frac{d - 1}{b} < 1$ y $\frac{d}{b} > -1$.

APLICACIONES

6.31 Se escoge un punto D en la línea AB, cuyo punto medio es C y cuya longitud es a. Si X, la distancia de D a A, es una variable aleatoria que tiene la densidad uniforme con $\alpha = 0$ y $\beta = a$, ¿cuál es la probabilidad de que AD, BD y AC formarán un triángulo?

6.32 En ciertos experimentos, el error cometido cuando se determina la densidad de una sustancia es una variable aleatoria que tiene la densidad uniforme con $\alpha = -0.015$ y $\beta = 0.015$. Encuentre las probabilidades de que un error tal

(a) estará entre -0.002 y 0.003;

(b) excederá 0.005 en valor absoluto.

6.33 Si una compañía emplea n vendedores, sus ventas brutas en miles de dólares se puede considerar como una variable aleatoria que tiene una distribución gamma con $\alpha = 80\sqrt{n}$ y $\beta = 2$. Si el costo de ventas es $8,000 dólares por vendedor, ¿cuántos vendedores debe emplear la compañía para maximizar la utilidad esperada?

6.34 En cierta ciudad, el consumo diario de energía eléctrica en millones de kilovatios-hora se puede tratar como una variable aleatoria que tiene la distribución gamma con $\alpha = 3$ y $\beta = 2$. Si la planta generadora de esta ciudad tiene una capacidad diaria de 12 millones de kilovatios-hora, ¿cuál es la probabilidad de que esta oferta de energía sea inadecuada cualquier día dado?

6.35 El millaje (en miles de millas) que los dueños de automóviles obtienen con una cierta marca de neumático radial es una variable aleatoria que tiene una distribución exponencial con $\theta = 40$. Encuentre las probabilidades de que uno de estos neumáticos durará

(a) al menos 20,000 millas;

(b) cuando mucho 30,000 millas.

6.36 La cantidad de tiempo que un reloj funcionará sin tener que volverlo a poner a tiempo es una variable aleatoria que tiene la distribución exponencial con $\theta = 120$ días. Encuentre las probabilidades de que un reloj tal

(a) tendrá que ponerse a tiempo en menos de 24 días;

(b) no tendrá que ponerse a tiempo en al menos 180 días.

6.37 El número de aviones que llegan por día a un pequeño aeropuerto privado es una variable aleatoria que tiene una distribución de Poisson con $\lambda = 28.8$. ¿Cuál es la probabilidad de que el tiempo entre dos llegadas tales sea al menos 1 hora?

6.38 El número total de cheques sin fondos que un banco recibe durante un día de negocios de 5 horas es una variable aleatoria de Poisson con $\lambda = 2$. ¿Cuál es la probabilidad de que no recibirá un cheque sin fondos en un día cualquiera durante las primeras 2 horas de actividad?

6.39 Una cierta clase de aparato doméstico requiere una reparación en promedio una vez cada 2 años. Suponiendo que los tiempos entre reparaciones están distribuidos exponencialmente, ¿cuál es la probabilidad de que un aparato doméstico tal trabajará al menos 3 años sin requerir reparaciones?

6.40 Si la proporción anual de declaraciones erróneas de impuestos al ingreso sometidas al IRS (*Internal Revenue Service*; servicio interno de recaudación) se puede considerar una variable aleatoria que tiene la distribución beta con $\alpha = 2$ y $\beta = 9$, ¿cuál es la probabilidad de que en un año dado cualquiera habrá menos del 10 por ciento de declaraciones erróneas?

6.41 Si la proporción anual de nuevos restaurantes que fracasan en una ciudad dada se puede considerar como una variable aleatoria que tiene una distribución beta con $\alpha = 1$ y $\beta = 4$, encuentre

(a) la media de esta distribución, esto es, la proporción anual de nuevos restaurantes que se puede esperar que fracasen en la ciudad dada;

(b) la probabilidad de que al menos 25 por ciento de todos los nuevos restaurantes fracasen en la ciudad dada en un año cualquiera.

6.42 Suponga que la vida de servicio de un semiconductor es una variable aleatoria que tiene la distribución de Weibull (véase el ejercicio 6.23) con $\alpha = 0.025$ y $\beta = 0.500$.

(a) ¿Cuánto se puede esperar que dure un semiconductor como ése?

(b) ¿Cuál es la probabilidad de que un semiconductor como ése estará todavía en condiciones operativas después de 4,000 horas?

6.5 LA DISTRIBUCIÓN NORMAL

La distribución normal, que estudiaremos en esta sección, es de muchas maneras, la piedra angular de la teoría estadística moderna. Se investigó por primera vez en el siglo XIX cuando los científicos observaron un grado asombroso de regularidad en los errores de medición. Encontraron que los patrones (distribuciones) que observaban se podían aproximar cercanamente por curvas continuas, a los que se referían como "curvas normales de errores" y las atribuían a las leyes del azar. Abraham de Moivre (1667-1745), Pierre Laplace (1749-1827) y Karl Gauss (1777-1855) estudiaron por primera vez las propiedades matemáticas de estas curvas normales.

DEFINICIÓN **6.6** Una variable aleatoria X tiene una **distribución normal** y se conoce como una variable aleatoria normal si y sólo si su densidad de probabilidad está dada por

$$n(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad \text{para } -\infty < x < \infty$$

donde $\sigma > 0$.

La gráfica de una distribución normal, con la forma de la sección transversal de una campana, se muestra en la figura 6.4.

La notación que aquí se usa es similar a la que se utiliza en relación con algunas de las distribuciones de probabilidad del capítulo 5: muestra explícitamente que los dos parámetros de la distribución normal son μ y σ. Sin embargo, queda por mostrar que el parámetro μ es, de hecho, $E(X)$ y que el parámetro σ es, de hecho, la raíz cuadrada

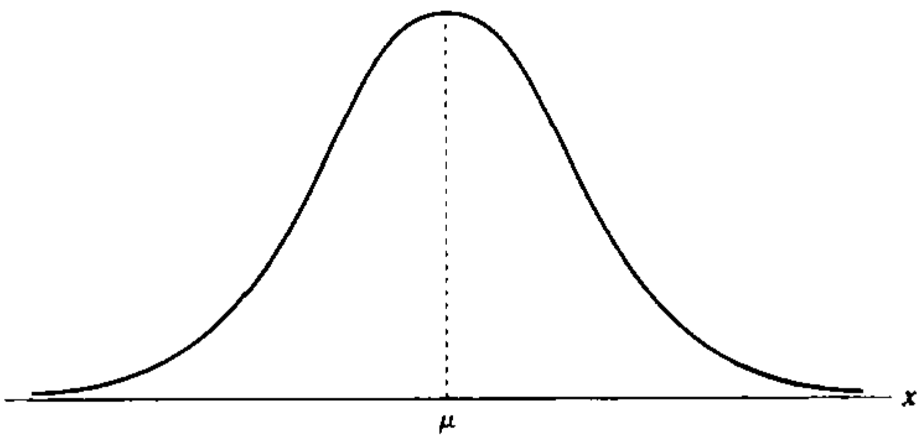

Figura 6.4 Gráfica de la distribución normal.

de var(X), donde X es una variable aleatoria que tiene la distribución normal con estos dos parámetros.

No obstante, primero demostremos que la fórmula de la definición 6.6 puede servir como una densidad de probabilidad. Puesto que los valores de $n(x; \mu, \sigma)$ son evidentemente positivos mientras $\sigma > 0$, debemos probar que el área total bajo la curva es igual a 1. Al integrar de $-\infty$ a ∞ y hacer la sustitución $z = \frac{x-\mu}{\sigma}$, obtenemos

$$\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2} dz = \frac{2}{\sqrt{2\pi}} \int_0^{\infty} e^{-\frac{1}{2}z^2} dz$$

Entonces, puesto que la integral de la derecha es igual a $\frac{\Gamma(\frac{1}{2})}{\sqrt{2}} = \frac{\sqrt{\pi}}{\sqrt{2}}$ de acuerdo al ejercicio 6.9, se sigue que el área total bajo la curva es igual a $\frac{2}{\sqrt{2\pi}} \cdot \frac{\sqrt{\pi}}{\sqrt{2}} = 1$.

A continuación probaremos que

TEOREMA 6.6 La función generatriz de momentos de la distribución normal está dada por

$$M_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$$

Demostración. Por definición,

$$\begin{aligned} M_X(t) &= \int_{-\infty}^{\infty} e^{xt} \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \\ &= \frac{1}{\sigma\sqrt{2\pi}} \cdot \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}[-2xt\sigma^2 + (x-\mu)^2]} dx \end{aligned}$$

y si completamos el cuadrado, esto es, usamos la identidad

$$-2xt\sigma^2 + (x-\mu)^2 = [x - (\mu + t\sigma^2)]^2 - 2\mu t\sigma^2 - t^2\sigma^4$$

obtenemos

$$M_X(t) = e^{\mu + \frac{1}{2}t^2\sigma^2}\left\{\frac{1}{\sigma\sqrt{2\pi}}\cdot\int_{-\infty}^{\infty} e^{-\frac{1}{2}\left[\frac{x-(\mu+t\sigma^2)}{\sigma}\right]^2}dx\right\}$$

Puesto que la cantidad dentro de las llaves es la integral de $-\infty$ a ∞ de una densidad normal con los parámetros $\mu + t\sigma^2$ y σ, y por tanto es igual a 1, se sigue que

$$M_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2} \qquad \blacktriangledown$$

Ahora estamos listos para verificar que los parámetros μ y σ en la definición 6.6 son, ciertamente, la media y la desviación estándar de la distribución normal. Si diferenciamos dos veces $M_X(t)$ con respecto a t, obtenemos

$$M'_X(t) = (\mu + \sigma^2 t)\cdot M_X(t)$$

y

$$M''_X(t) = \left[(\mu + \sigma^2 t)^2 + \sigma^2\right]\cdot M_X(t)$$

de manera que $M'_X(0) = \mu$ y $M''_X(0) = \mu^2 + \sigma^2$. Así, $E(X) = \mu$ y $\mathrm{var}(X) = (\mu^2 + \sigma^2) - \mu^2 = \sigma^2$.

Puesto que la distribución normal juega un papel básico en estadística y su densidad no se puede integrar directamente, se han tabulado sus áreas para el caso especial donde $\mu = 0$ y $\sigma = 1$.

> **DEFINICIÓN 6.7** La distribución normal con $\mu = 0$ y $\sigma = 1$ se conoce como la **distribución normal estándar**.

Los asientos en la tabla III, representados por el área sombreada de la figura 6.5, son los valores

$$\int_0^z \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}\,dx$$

0 z

Figura 6.5 Áreas tabuladas bajo la distribución normal estándar.

esto es, las probabilidades de que una variable aleatoria que tenga la distribución normal estándar asumirá un valor en el intervalo de 0 a z, para $z = 0.00, 0.01, 0.02, \ldots, 3.08$ y 3.09 y también $z = 4.0$, $z = 5.0$ y $z = 6.0$. En virtud de la simetría de la distribución normal alrededor de su media, es innecesario ampliar la tabla III a los valores negativos de z.

EJEMPLO 6.2

Encuentre las probabilidades de que una variable aleatoria que tiene la distribución normal estándar asumirá un valor

(a) menor que 1.72;

(b) menor que −0.88;

(c) entre 1.30 y 1.75;

(d) entre −0.25 y 0.45.

Solución

(a) Buscamos el asiento correspondiente a $z = 1.72$ en la tabla III, sumamos 0.500 (véase la figura 6.6) y obtenemos $0.4573 + 0.5000 = 0.9573$.

(b) Buscamos el asiento correspondiente a $z = 0.88$ en la tabla III, restamos 0.5000 (véase la figura 6.6) y obtenemos $0.5000 - 0.3106 = 0.1894$.

(c) Buscamos los asientos correspondientes a $z = 1.75$ y $z = 1.30$ en la tabla III, restamos el segundo del primero (véase la figura 6.6) y obtenemos $0.4599 - 0.4032 = 0.0567$.

(d) Buscamos los asientos correspondientes a $z = 0.25$ y $z = 0.45$ en la tabla III, los sumamos (véase la figura 6.6) y obtenemos $0.0987 + 0.1736 = 0.2723$. ▲

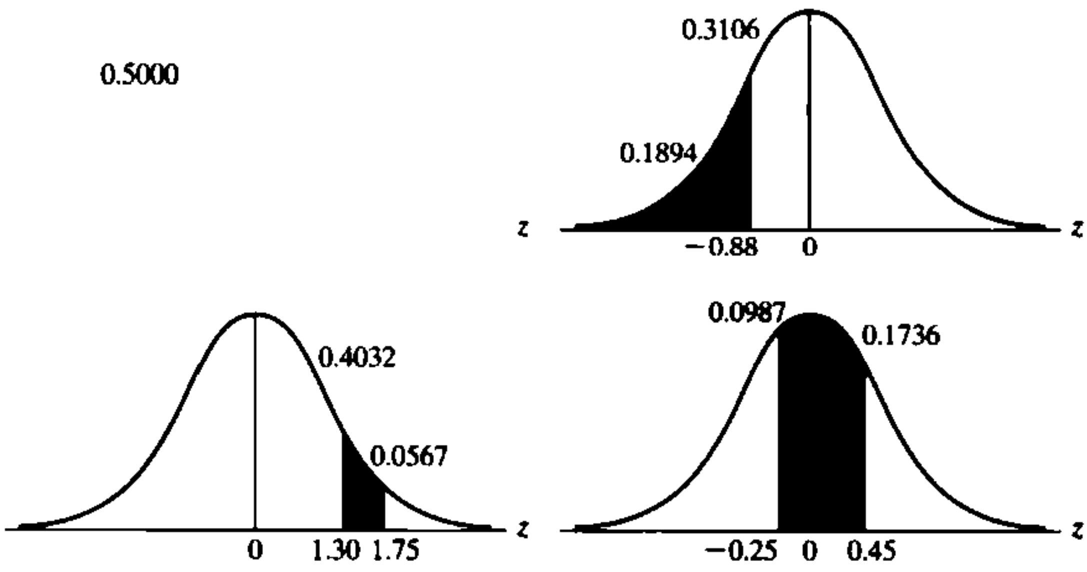

Figura 6.6 Diagramas para el ejemplo 6.2.

Ocasionalmente, se nos pide encontrar un valor de z que corresponda a una probabilidad especificada que cae entre los valores listados en la tabla III. En ese caso, por conveniencia, siempre escogemos el valor de z que corresponda al valor tabular que es más cercano a la probabilidad especificada. Sin embargo, si la probabilidad dada cae a mitad de camino entre los valores tabulares, escogeremos para z el valor que cae a mitad de camino entre los valores correspondientes de z.

EJEMPLO 6.3

Con respecto a la tabla III, encuentre los valores de z que corresponden a los datos de

(a) 0.3512;
(b) 0.2533.

Solución

(a) Puesto que 0.3512 cae entre 0.3508 y 0.3531, que corresponden a $z = 1.04$ y $z = 1.05$, y puesto que 0.3512 es más cercano a 0.3508 que 0.3531, escogemos $z = 1.04$.
(b) Puesto que 0.2533 cae a mitad de camino entre 0.2517 y 0.2549, que corresponden a $z = 0.68$ y $z = 0.69$, escogemos $z = 0.685$. ▲

Para determinar las probabilidades relacionadas con variables aleatorias que tienen distribuciones normales distintas a la distribución normal estándar, usamos el siguiente teorema

> TEOREMA 6.7 Si X tiene una distribución normal con media μ y la desviación estándar σ, entonces
>
> $$Z = \frac{X - \mu}{\sigma}$$
>
> tiene la distribución normal estándar.

Demostración. Puesto que la relación entre los valores de X y Z es lineal, Z debe asumir un valor entre $z_1 = \frac{x_1 - \mu}{\sigma}$ y $z_2 = \frac{x_2 - \mu}{\sigma}$ cuando X asume un valor entre x_1 y x_2. Por tanto podemos escribir

$$\begin{aligned}
P(x_1 < X < x_2) &= \frac{1}{\sqrt{2\pi}\sigma} \int_{x_1}^{x_2} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{z_1}^{z_2} e^{-\frac{1}{2}z^2} dz \\
&= \int_{z_1}^{z_2} n(z; 0, 1)\, dz \\
&= P(z_1 < Z < z_2)
\end{aligned}$$

donde se ve que Z es una variable aleatoria que tiene la distribución normal estándar. ▼

Así, para usar la tabla III en relación con cualquier variable aleatoria que tiene una distribución normal, simplemente efectuamos el cambio de escala $z = \frac{x - \mu}{\sigma}$.

EJEMPLO 6.4

Supongamos que la cantidad de radiación cósmica a la que una persona está expuesta cuando vuela en jet por Estados Unidos es una variable aleatoria que tiene una distribución normal con una media de 4.35 mrem y una desviación estándar de 0.59 mrem. ¿Cuál es la probabilidad de que una persona estará expuesta a más de 5.20 mrem de radiación cósmica en un vuelo como ése?

Solución

Buscamos el dato correspondiente a $z = \frac{5.20 - 4.35}{0.59} = 1.44$ en la tabla III y lo restamos de 0.5000 (véase la figura 6.7) y obtenemos $0.5000 - 0.4251 = 0.0749$. ▲

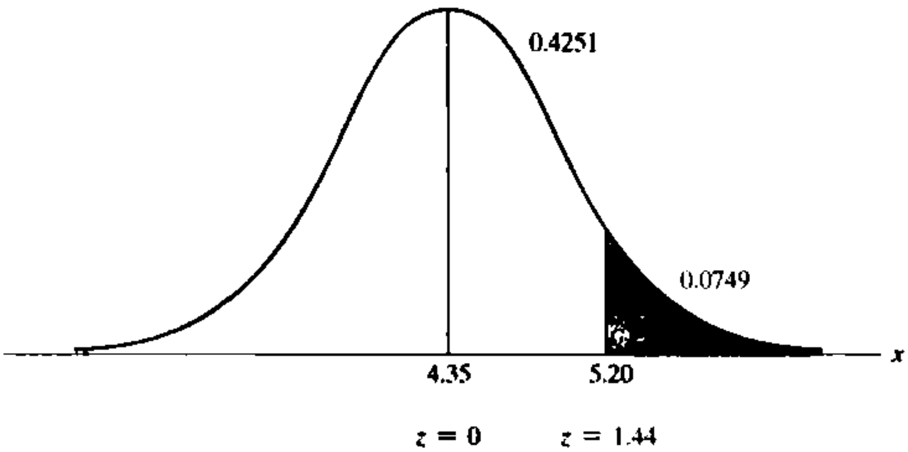

Figure 6.7 Diagrama para el ejemplo 6.4.

Con la ayuda de programas de computadora escritos especialmente para aplicaciones estadísticas es posible encontrar directamente las probabilidades relacionadas con variables aleatorias que tienen una distribución normal y otras distribuciones continuas más. El siguiente ejemplo ilustra dichos cálculos mediante el uso del software estadístico *MINITAB*.

EJEMPLO 6.5

Use un programa de computadora para encontrar la probabilidad de que una variable aleatoria que tiene

(a) la distribución ji cuadrada con 25 grados de libertad asumirá un valor mayor que 30;

(b) la distribución normal con la media 18.7 y la desviación estándar 9.1 asumirá un valor entre 10.6 y 24.8.

Solución

Mediante el uso del software *MINITAB*, seleccionamos la opción "distribución cumulativa" para obtener lo siguiente:

(a)
```
MTB>CDF C1;
SUBC>Chisquare 25
     30.0000   0.7757
```

Así, la probabilidad requerida es $1.0000 - 0.7757 = 0.2243$.

(b)
```
MTB>CDF C2;                    y      MTB>CDF C3;
SUBC>Normal 18.7 9.1.                 SUBC>Normal 18.7 9.1.
     10.6000   0.1867                      24.800   0.7487
```

Así, la probabilidad requerida es $0.7487 - 0.1867 = 0.5620$. ▲

6.6 LA APROXIMACIÓN NORMAL A LA DISTRIBUCIÓN BINOMIAL

Algunas veces se introduce la distribución normal como una distribución continua que proporciona una aproximación cercana a la distribución binomial cuando n, el número de ensayos, es muy grande y θ, la probabilidad de un éxito en un ensayo individual, es cercano a $\frac{1}{2}$. La figura 6.8 muestra los histogramas de las distribuciones binomiales con

$n = 2$ $n = 5$

$n = 10$ $n = 25$

Figure 6.8 Distribuciones binomiales con $\theta = \frac{1}{2}$.

$\theta = \frac{1}{2}$ y $n = 2, 5, 10$ y 25, y se puede ver que con n creciente estas distribuciones se aproximan al patrón simétrico en forma de campana de la distribución normal.

A fin de ofrecer un fundamento teórico para este argumento, probemos primero el siguiente teorema.

TEOREMA 6.8 Si X es una variable aleatoria que tiene una distribución binomial con los parámetros n y θ, entonces la función generatriz de momentos de

$$Z = \frac{X - n\theta}{\sqrt{n\theta(1-\theta)}}$$

se aproxima a la distribución normal estándar cuando $n \to \infty$.

Demostración. Al valernos de los teoremas 4.10 y 5.4, podemos escribir

$$M_Z(t) = M_{\frac{X-\mu}{\sigma}}(t) = e^{-\mu t/\sigma} \cdot [1 + \theta(e^{t/\sigma} - 1)]^n$$

donde $\mu = n\theta$ y $\sigma = \sqrt{n\theta(1-\theta)}$. Entonces, si tomamos logaritmos y sustituimos en la serie de Maclaurin de $e^{t/\sigma}$, obtenemos

$$\begin{aligned}
\ln M_{\frac{X-\mu}{\sigma}}(t) &= -\frac{\mu t}{\sigma} + n \cdot \ln[1 + \theta(e^{t/\sigma} - 1)] \\
&= -\frac{\mu t}{\sigma} + n \cdot \ln\left[1 + \theta\left\{\frac{t}{\sigma} + \frac{1}{2}\left(\frac{t}{\sigma}\right)^2 + \frac{1}{6}\left(\frac{t}{\sigma}\right)^3 + \cdots\right\}\right]
\end{aligned}$$

y, si usamos la serie infinita $\ln(1+x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \cdots$, la cual converge para $|x| < 1$, para expandir este logaritmo, se sigue que

$$\begin{aligned}
\ln M_{\frac{X-\mu}{\sigma}}(t) = &-\frac{\mu t}{\sigma} + n\theta\left[\frac{t}{\sigma} + \frac{1}{2}\left(\frac{t}{\sigma}\right)^2 + \frac{1}{6}\left(\frac{t}{\sigma}\right)^3 + \cdots\right] \\
&- \frac{n\theta^2}{2}\left[\frac{t}{\sigma} + \frac{1}{2}\left(\frac{t}{\sigma}\right)^2 + \frac{1}{6}\left(\frac{t}{\sigma}\right)^3 + \cdots\right]^2 \\
&+ \frac{n\theta^3}{3}\left[\frac{t}{\sigma} + \frac{1}{2}\left(\frac{t}{\sigma}\right)^2 + \frac{1}{6}\left(\frac{t}{\sigma}\right)^3 + \cdots\right]^3 - \cdots
\end{aligned}$$

Al reunir las potencias de t, obtenemos

$$\begin{aligned}
\ln M_{\frac{X-\mu}{\sigma}}(t) &= \left(-\frac{\mu}{\sigma} + \frac{n\theta}{\sigma}\right)t + \left(\frac{n\theta}{2\sigma^2} - \frac{n\theta^2}{2\sigma^2}\right)t^2 \\
&\quad + \left(\frac{n\theta}{6\sigma^3} - \frac{n\theta^2}{2\sigma^3} + \frac{n\theta^3}{3\sigma^3}\right)t^3 + \cdots \\
&= \frac{1}{\sigma^2}\left(\frac{n\theta - n\theta^2}{2}\right)t^2 + \frac{n}{\sigma^3}\left(\frac{\theta - 3\theta^2 + 2\theta^3}{6}\right)t^3 + \cdots
\end{aligned}$$

puesto que $\mu = n\theta$. Entonces, al sustituir $\sigma = \sqrt{n\theta(1 - \theta)}$, encontramos que

$$\ln M_{\frac{X-\mu}{\sigma}}(t) = \frac{1}{2}t^2 + \frac{n}{\sigma^3}\left(\frac{\theta - 3\theta^2 + 2\theta^3}{6}\right)t^3 + \cdots$$

donde para $r > 2$ el coeficiente de t^r es una constante multiplicada por $\frac{n}{\sigma^r}$, la cual se aproxima a 0 cuando $n \to \infty$. Se sigue que

$$\lim_{n\to\infty} \ln M_{\frac{X-\mu}{\sigma}}(t) = \frac{1}{2}t^2$$

y puesto que el límite de un logaritmo es igual al logaritmo del límite (siempre que los dos límites existan), concluimos que

$$\lim_{n\to\infty} M_{\frac{X-\mu}{\sigma}}(t) = e^{\frac{1}{2}t^2}$$

la cual es la función generatriz de momentos del teorema 6.6 con $\mu = 0$ y $\sigma = 1$ ▼

Esto completa la demostración del teorema 6.8, pero ¿ya demostramos que cuando $n \to \infty$ la distribución de Z, la variable aleatoria binomial **estandarizada**, se aproxima a la distribución normal estándar? No exactamente. Para este fin, debemos referirnos a dos teoremas que enunciaremos sin demostrarlos:

1. *Hay una correspondencia unívoca entre las funciones generatrices de momentos y las distribuciones (densidades) de probabilidad cuando existen las primeras*
2. *Si la función generatriz de momentos de una variable aleatoria se aproxima a la de otra variable aleatoria, entonces la distribución (densidad) de la primera variable aleatoria se aproxima a la de la segunda variable aleatoria bajo las mismas condiciones límite.*

Hablando estrictamente, nuestros resultados son válidos cuando $n \to \infty$, pero a menudo se usa la distribución normal para aproximar probabilidades binomiales aun cuando n es relativamente pequeña. Una buena regla empírica es usar esta aproximación sólo cuando $n\theta$ y $n(1 - \theta)$ son ambos mayores que 5.

EJEMPLO 6.6

Use la aproximación normal a la distribución binomial para determinar la probabilidad de sacar seis caras y 10 cruces en 16 lanzamientos de una moneda balanceada o equilibrada.

Solución

Para encontrar esta aproximación debemos usar la **corrección de continuidad**, de acuerdo a la cual cada entero no negativo k se representa con el intervalo de $k - \frac{1}{2}$ a $k + \frac{1}{2}$. Con respecto a la figura 6.9, debemos determinar así el área bajo la curva entre 5.5 y 6.5, y puesto que $\mu = 16 \cdot \frac{1}{2} = 8$ y $\sigma = \sqrt{16 \cdot \frac{1}{2} \cdot \frac{1}{2}} = 2$, debemos encontrar el área entre:

$$z = \frac{5.5 - 8}{2} = -1.25 \quad \text{y} \quad z = \frac{6.5 - 8}{2} = -0.75$$

Los datos en la tabla III que corresponden a $z = 1.25$ y $z = 0.75$ son 0.3944 y 0.2734, y encontramos que la aproximación normal a la probabilidad de "seis caras y 10 cruces" es $0.3944 - 0.2734 = 0.1210$. Puesto que el valor correspondiente en la tabla I es 0.1222, encontramos que el error de la aproximación es -0.0012 y que el porcentaje de error es $\frac{0.0012}{0.1222} \cdot 100 = 0.98\%$ en valor absoluto. ▲

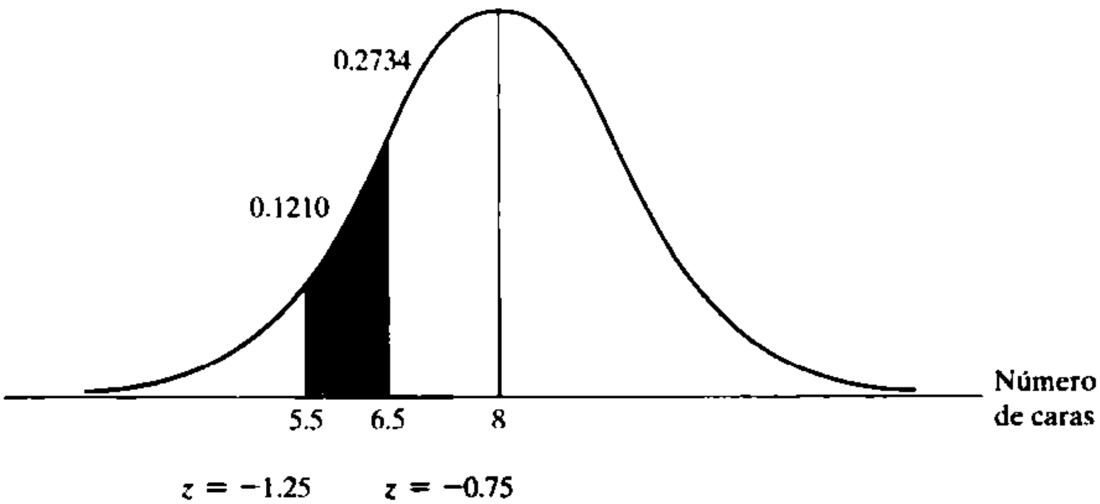

Figura 6.9 Diagrama para el ejemplo 6.6.

La aproximación normal a la distribución binomial solía aplicarse muy extensamente, particularmente al aproximar probabilidades relacionadas con grandes valores de variables aleatorias binomiales. Hoy en día, la mayor parte de este trabajo se hace con computadoras, como se ilustra en el ejemplo 6.5, y hemos mencionado la relación entre las distribuciones normal y binomial principalmente a causa de sus aplicaciones teóricas. Constituye la base de muchos de los procedimientos tratados en los capítulos 11, 13 y 16.

EJERCICIOS

6.43 Demuestre que la distribución normal tiene

(a) un máximo relativo en $x = \mu$;

(b) puntos de inflexión en $x = \mu - \sigma$ y $x = \mu + \sigma$.

6.44 Muestre que la ecuación diferencial del ejercicio 6.30 con $b = c = 0$ y $a > 0$ nos da una distribución normal.

6.45 En la demostración del teorema 6.6 diferenciamos dos veces la función generatriz de momentos de la distribución normal con respecto a t para demostrar que $E(X) = \mu$ y $\text{var}(X) = \sigma^2$. Al diferenciar dos veces más y usar la fórmula del ejercicio 4.33, encuentre las expresiones para μ_3 y μ_4.

6.46 Si X es una variable aleatoria que tiene una distribución normal con la media μ y la desviación estándar σ, encuentre la función generatriz de momentos de $Y = X - c$, donde c es una constante, y úsela para volver a trabajar con el ejercicio 6.45.

6.47 Use los resultados del ejercicio 6.45 para demostrar que $\alpha_3 = 0$ y $\alpha_4 = 3$ para distribuciones normales, donde α_3 y α_4 son como se definieron en los ejercicios 4.34 y 4.35.

6.48 Si X es una variable aleatoria que tiene una distribución normal con la media μ y la desviación estándar σ, use la tercera parte del teorema 4.10 y el teorema 6.6 para demostrar que la función generatriz de momentos de,

$$Z = \frac{X - \mu}{\sigma}$$

es la función generatriz de momentos de la distribución normal estándar. Advierta que, junto con los dos teoremas en la página 224, esto demuestra el teorema 6.7.

6.49 Si X es una variable aleatoria que tiene la distribución normal estándar y $Y = X^2$, demuestre que $\text{cov}(X, Y) = 0$ aun cuando X y Y no son evidentemente independientes.

6.50 Use la expansión en la serie de Maclaurin de la función generatriz de momentos de la distribución normal estándar para demostrar que

(a) $\mu_r = 0$ cuando r es impar;

(b) $\mu_r = \dfrac{r!}{2^{r/2}\left(\dfrac{r}{2}\right)!}$ cuando r es par.

6.51 Si hacemos $K_X(t) = \ln M_{X-\mu}(t)$, el coeficiente de $\dfrac{t^r}{r!}$ en la serie de Maclaurin de $K_X(t)$ se llama el **résimo acumulante**, y se denota con κ_r. Al igualar coeficientes de potenciales iguales, demuestre que

(a) $\kappa_2 = \mu_2$;

(b) $\kappa_3 = \mu_3$;

(c) $\kappa_4 = \mu_4 - 3\mu_2^2$.

6.52 Con respecto al ejercicio 6.51, muestre que para las distribuciones normales $\kappa_2 = \sigma^2$ todos los demás acumulantes son cero.

6.53 Pruebe que si X es variable aleatoria que tiene la distribución de Poisson con el parámetro λ y $\lambda \to \infty$, entonces la función generatriz de momentos de

$$Z = \frac{X - \lambda}{\sqrt{\lambda}}$$

esto es, la de la variable aleatoria de Poisson estandarizada, se aproxima a la función generatriz de momentos de la distribución normal estándar.

6.54 Demuestre que cuando $\alpha \to \infty$ y β permanece constante, la función generatriz de momentos de una variable aleatoria gamma estandarizada se aproxima a la función generatriz de momentos de la distribución normal estándar.

APLICACIONES

6.55 Si Z es una variable aleatoria que tiene la distribución normal estándar, encuentre las probabilidades que asumirá un valor

(a) mayor que 1.14;
(b) mayor que -0.36;
(c) entre -0.46 y -0.09;
(d) entre -0.58 y 1.12.

6.56 Si Z es una variable aleatoria que tiene la distribución normal estándar, encuentre

(a) $P(Z < 1.33)$;
(b) $P(Z \leq -0.79)$;
(c) $P(0.55 < Z < 1.22)$;
(d) $P(-1.90 \leq Z \leq 0.44)$.

6.57 Encuentre z si el área de la curva normal estándar

(a) entre 0 y z es 0.4726;
(b) a la izquierda de z es 0.9868;
(c) a la derecha de z es 0.1314;
(d) entre $-z$ y z es 0.8502.

6.58 Si Z es una variable aleatoria que tiene la distribución normal estándar, encuentre los valores respectivos de z_1, z_2, z_3 y z_4 tales que

(a) $P(0 < Z < z_1) = 0.4306$;
(b) $P(Z \geq z_2) = 0.7704$;
(c) $P(Z > z_3) = 0.2912$;
(d) $P(-z_4 \leq Z < z_4) = 0.9700$.

6.59 Si X es una variable aleatoria que tiene una distribución normal, ¿cuáles son las probabilidades de obtener un valor

(a) dentro de una desviación estándar de la media;
(b) dentro de dos desviaciones estándar de la media;
(c) dentro de tres desviaciones estándar de la media;
(d) dentro de cuatro desviaciones estándar de la media?

6.60 Si z_α está definida por

$$\int_{z_\alpha}^{\infty} n(z; 0, 1)\, dz = \alpha$$

encuentre sus valores para

(a) $\alpha = 0.05$;
(b) $\alpha = 0.025$;
(c) $\alpha = 0.01$;
(d) $\alpha = 0.005$.

6.61 (a) Use un programa de computadora para encontrar la probabilidad de que una variable aleatoria que tiene la distribución normal con la media -1.786 y la desviación estándar 1.0416 asumirá un valor entre -2.159 y 0.5670.

(b) Interpole en la tabla III para encontrar esta probabilidad y compare su resultado con el valor más exacto encontrado en el inciso (a).

6.62 (a) Use un programa de computadora para encontrar la probabilidad de que una variable aleatoria que tiene la distribución normal con media 5.853 y la desviación estándar 1.361 asumirá un valor mayor que 8.625;

(b) Interpole en la tabla III para encontrar esta probabilidad y compare su resultado con el valor más exacto encontrado en el inciso (a).

6.63 Suponga que durante los periodos de meditación trascendental la reducción del consumo de oxígeno de una persona es una variable aleatoria que tiene una distribución normal con $\mu = 37.6$ cc por minuto y $\sigma = 4.6$ cc por minuto. Encuentre las probabilidades de que durante un periodo de meditación trascendental el consumo de oxígeno de una persona se reducirá por

(a) al menos 44.5 cc por minuto;

(b) cuando mucho 35.0 cc por minuto;

(c) cualquier valor entre 30.0 y 40.0 cc por minuto.

6.64 En un proceso fotográfico, el tiempo de revelado de impresiones se puede considerar como una variable aleatoria que tiene la distribución normal con $\mu = 15.40$ segundos y $\sigma = 0.48$ segundos. Encuentre las probabilidades de que el tiempo que toma revelar una de las impresiones será

(a) al menos 16.00 segundos;

(b) cuando mucho 14.20 segundos;

(c) cualquier valor entre 15.00 y 15.80 segundos.

6.65 Supongamos que la cantidad real de café instantáneo que una máquina sirve en un frasco de "6 onzas" es una variable aleatoria que tiene una distribución normal con $\sigma = 0.05$ onzas. Si sólo 3 por ciento de los frascos deben contener menos de 6 onzas de café, ¿cuál debe ser la media del llenado de estos frascos?

6.66 Una variable aleatoria tiene la distribución normal con $\sigma = 10$. Si la probabilidad de que la variable aleatoria asumirá un valor menor que 82.5 es 0.8212, ¿cuál es la probabilidad de que asumirá un valor mayor que 58.3?

6.67 Verifique en cada caso si la aproximación normal a la distribución binomial se puede usar de acuerdo a la regla empírica de la página 224.

(a) $n = 16$ y $\theta = 0.20$;

(b) $n = 65$ y $\theta = 0.10$;

(c) $n = 120$ y $\theta = 0.98$.

6.68 Suponga que queremos usar la aproximación normal a la distribución binomial para determinar $b(1; 150, 0.05)$.

(a) Con base en la regla empírica de la página 224, ¿estaría justificado el uso de la aproximación?

(b) Haga la aproximación y redondee a cuatro decimales.

(c) Si una impresión de computadora muestra que $b(1; 150, 0.05) = 0.0036$ redondeado a cuatro decimales, ¿cuál es el porcentaje de error de la aproximación obtenida en el inciso (b)?

Esto sirve para ilustrar que la regla empírica es sólo eso y no más; hacer aproximaciones como ésta también requiere una buena cantidad de juicio profesional.

6.69 Con respecto al ejercicio 6.68, demuestre que la distribución de Poisson hubiera dado una mejor aproximación.

6.70 Use la aproximación normal a la distribución binomial para determinar (con cuatro decimales) la probabilidad de obtener siete caras y siete cruces en 14 lanzamientos de una moneda balanceada. También consulte la tabla I para encontrar el error de esta aproximación.

6.71 Si 23 por ciento de todos los pacientes con presión arterial alta tiene efectos colaterales perjudiciales con cierta clase de medicamento, use la aproximación normal para encontrar la probabilidad de que entre 120 pacientes con presión arterial alta tratados con este medicamento más de 32 tendrán efectos colaterales perjudiciales.

6.72 Si la probabilidad de que un banco rechazará una solicitud de préstamo es 0.20, use la aproximación normal para determinar (con tres decimales) la probabilidad de que el banco rechazará cuando mucho 40 de 225 solicitudes de préstamo.

6.73 Para ilustrar la ley de los grandes números (véase también el ejercicio 5.30), use la aproximación normal a la distribución binomial para determinar las probabilidades de que la proporción de caras estará entre 0.49 y 0.51 cuando se lanza una moneda balanceada

(a) 100 veces;

(b) 1,000 veces;

(c) 10,000 veces.

6.7 LA DISTRIBUCIÓN NORMAL BIVARIADA

Entre las densidades multivariadas, la **distribución normal multivariada** es de especial importancia, la cual es una generalización de la distribución normal en una variable. Como es mejor (en realidad, es prácticamente necesario) presentar esta distribución en notación matricial, aquí sólo daremos el caso **bivariado**; los análisis del caso más general se encuentran entre las referencias al final de este capítulo.

DEFINICIÓN 6.8 Un par de variables aleatorias X y Y tienen una **distribución normal bivariada** y se conocen como variables aleatorias distribuidas normalmente en forma conjunta si y sólo si su densidad de probabilidad conjunta está dada por

$$f(x, y) = \frac{e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}$$

para $-\infty < x < \infty$ y $-\infty < y < \infty$, donde $\sigma_1 > 0$, $\sigma_2 > 0$, y $-1 < \rho < 1$.

Para estudiar esta distribución conjunta, mostremos primero que los parámetros μ_1, μ_2, σ_1 y σ_2 son las medias y las desviaciones estándar de las dos variables aleatorias X y Y. Para empezar, integramos sobre y de $-\infty$ a ∞ y obtenemos

$$g(x) = \frac{e^{-\frac{1}{2(1-\rho^2)}\left(\frac{x-\mu_1}{\sigma_1}\right)^2}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{y-\mu_2}{\sigma_2}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right)\right]} dy$$

para la densidad marginal de X. Entonces, hacemos la sustitución temporal $u = \dfrac{x-\mu_1}{\sigma_1}$ para simplificar la notación y si cambiamos la variable de integración al hacer $v = \dfrac{y-\mu_2}{\sigma_2}$, obtenemos

$$g(x) = \frac{e^{-\frac{1}{2(1-\rho^2)}u^2}}{2\pi\sigma_1\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2(1-\rho^2)}(v^2-2\rho uv)} dv$$

Después de completar el cuadrado al hacer

$$v^2 - 2\rho uv = (v - \rho u)^2 - \rho^2 u^2$$

y reunir términos, esto se vuelve

$$g(x) = \frac{e^{-\frac{1}{2}u^2}}{\sigma_1\sqrt{2\pi}} \left\{ \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{v-\rho u}{\sqrt{1-\rho^2}}\right)^2} dv \right\}$$

Finalmente, identificamos la cantidad entre paréntesis como la integral de una densidad normal de $-\infty$ a ∞ y, por tanto, al hacer igual a 1, obtenemos

$$g(x) = \frac{e^{-\frac{1}{2}u^2}}{\sigma_1\sqrt{2\pi}} = \frac{1}{\sigma_1\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2}$$

para $-\infty < x < \infty$. Se sigue por inspección que la densidad marginal de X es una distribución normal con la media μ_1 y la desviación estándar σ_1 y, por simetría, que la densidad marginal de Y es una distribución normal con la media μ_2 y la desviación estándar σ_2.

En lo tocante al parámetro ρ donde ρ es la minúscula de la letra griega *rho*, se llama el **coeficiente de correlación**, y la integración necesaria mostrará que $\text{cov}(X, Y) = \rho\sigma_1\sigma_2$. Así, el parámetro ρ mide cómo varían juntas las variables aleatorias X y Y, y su significado se analizará con más detalle en el capítulo 14.

Cuando tratamos con un par de variables aleatorias que tienen una distribución normal bivariada, sus densidades condicionales también son de importancia; demostremos el siguiente teorema.

TEOREMA 6.9 Si X y Y tienen una distribución normal bivariada, la densidad condicional de Y dado $X = x$ es una distribución normal con la media

$$\mu_{Y|x} = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x - \mu_1)$$

y la varianza

$$\sigma^2_{Y|x} = \sigma^2_2(1 - \rho^2)$$

y la densidad condicional de X dado $Y = y$ es una distribución normal con la media

$$\mu_{X|y} = \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(y - \mu_2)$$

y la varianza

$$\sigma^2_{X|y} = \sigma^2_1(1 - \rho^2)$$

Demostración. Si escribimos $w(y|x) = \dfrac{f(x, y)}{g(x)}$ de acuerdo a la definición 3.13 y hacemos $u = \dfrac{x - \mu_1}{\sigma_1}$ y $v = \dfrac{y - \mu_2}{\sigma_2}$ para simplificar la notación, obtenemos

$$w(y|x) = \frac{\dfrac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\, e^{-\frac{1}{2(1-\rho^2)}[u^2 - 2\rho uv + v^2]}}{\dfrac{1}{\sqrt{2\pi}\sigma_1}\, e^{-\frac{1}{2}u^2}}$$

$$= \frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}}\, e^{-\frac{1}{2(1-\rho^2)}[v^2 - 2\rho uv + \rho^2u^2]}$$

$$= \frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}}\, e^{-\frac{1}{2}\left[\frac{v-\rho u}{\sqrt{1-\rho^2}}\right]^2}$$

Entonces, expresamos este resultado en términos de las variables originales, y obtenemos

$$w(y|x) = \frac{1}{\sigma_2\sqrt{2\pi}\sqrt{1-\rho^2}}\, e^{-\frac{1}{2}\left[\frac{y-\left\{\mu_2+\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1)\right\}}{\sigma_2\sqrt{1-\rho^2}}\right]^2}$$

para $-\infty < y < \infty$, y por inspección se puede ver que ésta es una distribución normal con la media $\mu_{Y|x} = \mu_2 + \rho\dfrac{\sigma_2}{\sigma_1}(x - \mu_1)$ y la varianza $\sigma^2_{Y|x} = \sigma^2_2(1 - \rho^2)$.

Los resultados correspondientes para la densidad condicional de X dado $Y = y$ siguen por simetría. ▼

La distribución normal bivariada tiene muchas propiedades importantes, algunas estadísticas y algunas puramente matemáticas. Entre las primeras, está la siguiente propiedad, que se pedirá al lector la demuestre en el ejercicio 6.74.

> **TEOREMA 6.10** Si dos variables aleatorias tienen una distribución normal bivariada, son independientes si y sólo si $\rho = 0$.

En relación con esto, si $\rho = 0$, se dice que las variables aleatorias no están **correlacionadas**.

También, hemos mostrado que para dos variables aleatorias que tienen una distribución normal bivariada las dos densidades marginales son normales, pero la recíproca no es necesariamente verdad. En otras palabras, las distribuciones marginales pueden ser normales ambas sin que la distribución conjunta sea una distribución normal bivariada. Por ejemplo, si la densidad bivariada de X y Y está dada por

$$f^*(x, y) = \begin{cases} 2f(x, y) & \text{dentro de los cuadros 2 y 4 de la figura 6.10} \\ 0 & \text{dentro de los cuadros 1 y 2 de la figura 6.10} \\ f(x, y) & \text{en cualquier otra parte} \end{cases}$$

donde $f(x, y)$ es el valor de la densidad normal bivariada con $\mu_1 = 0$, $\mu_2 = 0$ y $\rho = 0$ en (x, y), es fácil ver que las densidades marginales de X y Y son normales aun cuando su densidad conjunta no es una distribución normal bivariada.

Se obtienen muchas propiedades interesantes de la densidad normal bivariada al estudiar la **superficie normal bivariada**, que se muestra en la figura 6.11 y cuya ecuación es $z = f(x, y)$, donde $f(x, y)$ es el valor de la densidad normal bivariada en (x, y). Como se pedirá al lector que verifique en los ejercicios que siguen, la densidad normal

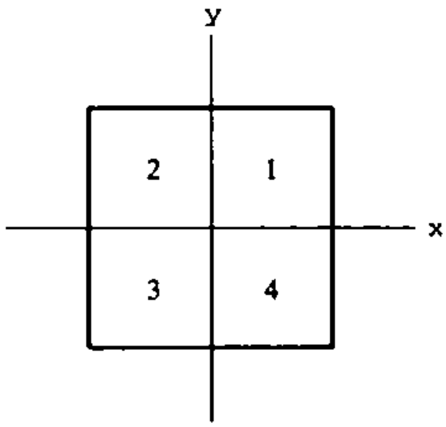

Figura 6.10 Espacio muestral para la densidad bivariada dada por $f^*(x, y)$.

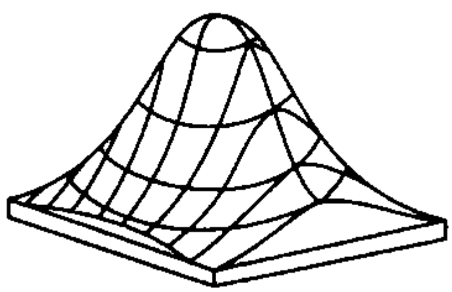

Figura 6.11 Superficie normal bivariada.

bivariada tiene un máximo en (μ_1, μ_2), cualquier plano paralelo al eje z interseca la superficie en una curva que tiene la forma de una distribución normal y cualquier plano paralelo al plano xy que interseca la superficie la interseca en una elipse llamada un **contorno de densidad de probabilidad constante**. Cuando $\rho = 0$ y $\sigma_1 = \sigma_2$, los contornos de densidad de probabilidad constante son círculos, y es costumbre referirse a la densidad conjunta correspondiente como una **distribución normal circular.**

EJERCICIOS

6.74 Para probar el teorema 6.10, demuestre que si X y Y tienen una distribución normal bivariada, entonces

(a) su independencia implica que $\rho = 0$;

(b) $\rho = 0$ implica que son independientes.

6.75 Demuestre que cualquier plano perpendicular al plano xy interseca la superficie normal bivariada en una curva que tiene la forma de una distribución normal.

6.76 Si el exponente de e de una densidad normal bivariada es

$$\frac{-1}{102}\left[(x+2)^2 - 2.8(x+2)(y-1) + 4(y-1)^2\right]$$

encuentre

(a) $\mu_1, \mu_2, \sigma_1, \sigma_2$ y ρ;

(b) $\mu_{Y|x}$ y $\sigma^2_{Y|x}$.

6.77 Si el exponente de e de una densidad normal bivariada es

$$\frac{-1}{54}\left(x^2 + 4y^2 + 2xy + 2x + 8y + 4\right)$$

encuentre σ_1, σ_2 y ρ, dado que $\mu_1 = 0$ y $\mu_2 = -1$.

6.78 Si X y Y tienen la distribución normal bivariada con $\mu_1 = 2$, $\mu_2 = 5$, $\sigma_1 = 3$, $\sigma_2 = 6$ y $\rho = \frac{2}{3}$, encuentre $\mu_{Y|1}$ y $\sigma_{Y|1}$.

6.79 Si X y Y tienen una distribución normal bivariada y $U = X + Y$ y $V = X - Y$, encuentre una expresión para el coeficiente de correlación de U y V.

6.80 Si X y Y tienen una distribución normal bivariada, se puede demostrar que su función generatriz de momentos conjunta (véase el ejercicio 4.63) está dada por

$$\begin{aligned} M_{X,Y}(t_1, t_2) &= E\left(e^{t_1 X + t_2 Y}\right) \\ &= e^{t_1\mu_1 + t_2\mu_2 + \frac{1}{2}(\sigma_1^2 t_1^2 + 2\rho\sigma_1\sigma_2 t_1 t_2 + \sigma_2^2 t_2^2)} \end{aligned}$$

Verifique que

(a) la primera derivada parcial de esta función con respecto a t_1 en $t_1 = 0$ y $t_2 = 0$ es μ_1;

(b) la segunda derivada parcial con respecto a t_1 en $t_1 = 0$ y $t_2 = 0$ es $\sigma_1^2 + \mu_1^2$;

(c) la segunda derivada parcial con respecto a t_1 y t_2 en $t_1 = 0$ y $t_2 = 0$ es $\rho\sigma_1\sigma_2 + \mu_1\mu_2$.

APLICACIONES

6.81 El centro de un blanco se toma como el origen de un sistema rectangular de coordenadas, con respecto al cual el punto de impacto de un cohete tiene las coordenadas X y Y. Si X y Y tienen la densidad normal bivariada con $\mu_1 = 0$, $\mu_2 = 0$, $\sigma_1 = 120$ pies, $\sigma_2 = 120$ pies y $\rho = 0$, encuentre la probabilidad de que el punto de impacto estará

(a) dentro de un cuadro con lados de 180 pies, cuyo centro está en el origen y cuyos lados son paralelos a los ejes de coordenadas;

(b) dentro de un círculo cuyo radio de 75 pies con su centro en el origen.

6.82 Si X y Y tienen la distribución normal circular con $\mu_1 = \mu_2 = 0$ y $\sigma_1 = \sigma_2 = 12$, encuentre

(a) la probabilidad de obtener un punto (x, y) dentro de un círculo $x^2 + y^2 = 36$;

(b) el valor de c para el cual la probabilidad de obtener un punto (x, y) dentro del círculo $x^2 + y^2 = c^2$ es 0.80.

6.83 Suponga que X y Y, la altura y el peso de ciertos animales, tienen la distribución normal bivariada con $\mu_1 = 18$ pulgadas, $\mu_2 = 15$ libras, $\sigma_1 = 3$ pulgadas, $\sigma_2 = 2$ libras y $\rho = 0.75$. Encuentre

(a) el peso esperado de uno de estos animales de 17 pulgadas de alto;

(b) la altura esperada de uno de estos animales de 20 libras de peso.

REFERENCIAS

En forma resumida es posible encontrar información útil acerca de varias densidades de probabilidades especiales en

DERMAN, C., GLESER, L., and OLKIN, I., *Probability Models and Applications*. Nueva York, Macmillan Publishing Co., Inc., 1980,

HASTINGS, N. A. J., and PEACOCK, J. B., *Statistical Distributions*. Londres: Butterworth & Co. Ltd., 1975,

y

JOHNSON, N. L., and KOTZ, S., *Continuous Univariate Distributions*, Vols. 1 and 2. Boston: Houghton Mifflin Company, 1970.

Una prueba directa que la distribución binomial estandarizada se aproxima a la distribución normal estándar cuando $n \to \infty$ se da en

KEEPING, E. S., *Introduction to Statistical Inference*. Princeton, N.J.: D. Van Nostrand Co., Inc., 1962.

Un tratamiento detallado de las propiedades matemáticas y estadísticas de la superficie normal bivariada se puede encontrar en

YULE, G. U., and KENDALL, M. G., *An Introduction to the Theory of Statistics*, 14th ed. Nueva York: Hafner Publishing Co., Inc., 1950.

La distribución normal multivariada se trata en notación matricial en

BICKEL, P. J., and DOKSUM, K. A., *Mathematical Statistics: Basic Ideas and Selected Topics*. San Francisco: Holden-Day, Inc., 1977,

HOGG, R. V., and CRAIG, A. T., *Introduction to Mathematical Statistics*, 4th ed. Nueva York: Macmillan Publishing Co., Inc., 1978,

LINDGREN, B. W., *Statistical Theory*, 3rd ed. Nueva York: Macmillan Publishing Co., Inc., 1976.

CAPÍTULO

7

Funciones de variables aleatorias

7.1 INTRODUCCIÓN

En este capítulo trataremos el problema de encontrar las densidades o probabilidades de distribución de **funciones de una o más variables aleatorias**. Esto es, dado un conjunto de variables aleatorias $X_1, X_2, \ldots, X_n$ y su densidad o distribución de probabilidad conjunta, estaremos interesados en encontrar la densidad o distribución de probabilidad de alguna variable aleatoria $Y = u(X_1, X_2, \ldots, X_n)$. Esto significa que los valores de Y están relacionados a los de las X es por medio de la ecuación

$$y = u(x_1, x_2, \ldots, x_n).$$

Para resolver esta clase de problemas se dispone de varios métodos. Los que examinaremos en las cuatro secciones siguientes son llamados la **técnica de la función de distribución**, la **técnica de transformación**, y la **técnica de la función generatriz de momentos**. Aunque en algunas situaciones se pueden usar los tres métodos, en la mayoría de los problemas será preferible una técnica (más fácil que las demás). Esto es cierto, por ejemplo, en algunos casos donde la función en cuestión es lineal en las variables aleatorias $X_1, X_2, \ldots, X_n$, la técnica de la función generatriz de momentos produce las derivaciones más simples.

Las diversas técnicas que examinaremos en este capítulo se usarán otra vez en el capítulo 8 para derivar varias distribuciones que son de fundamental importancia en la inferencia estadística.

7.2 TÉCNICA DE LA FUNCIÓN DE DISTRIBUCIÓN

Un método directo de obtener la densidad de probabilidad de una función de variables aleatorias continuas consiste en encontrar primero su función de distribución y después su densidad de probabilidad por diferenciación. Así, si $X_1, X_2, \ldots, X_n$ son variables aleatorias continuas con una densidad de probabilidad conjunta, la densidad de probabilidad de $Y = u(X_1, X_2, \ldots, X_n)$ se obtiene al determinar primero una expresión para la probabilidad

$$F(y) = P(Y \leqq y) = P[u(X_1, X_2, \ldots, X_n) \leqq y]$$

y después se diferencia para obtener

$$f(y) = \frac{dF(y)}{dy}$$

de acuerdo al teorema 3.6.

EJEMPLO 7.1

Si la densidad de probabilidad de X está dada por

$$f(x) = \begin{cases} 6x(1 - x) & \text{para } 0 < x < 1 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

encuentre la densidad de probabilidad de $Y = X^3$.

Solución

Sea $G(y)$ el valor de la función de distribución de Y en y, podemos escribir

$$\begin{aligned} G(y) &= P(Y \leqq y) \\ &= P(X^3 \leqq y) \\ &= P(X \leqq y^{1/3}) \\ &= \int_0^{y^{1/3}} 6x(1 - x)\, dx \\ &= 3y^{2/3} - 2y \end{aligned}$$

y por tanto

$$g(y) = 2(y^{-1/3} - 1)$$

para $0 < y < 1$; en cualquier otra parte, $g(y) = 0$. En el ejercicio 7.20 se le pedirá el lector que verifique este resultado mediante una técnica diferente. ▲

EJEMPLO 7.2

Si $Y = |X|$, demostrar que

$$g(y) = \begin{cases} f(y) + f(-y) & \text{para } y > 0 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

donde $f(x)$ es el valor de la densidad de probabilidad de X en x y $g(y)$es el valor de densidad de probabilidad de Y en y. También, use este resultado para encontrar la densidad de probabilidad de $Y = |X|$ cuando X tiene la distribución normal estándar.

Solución

Para $y > 0$ tenemos

$$\begin{aligned} G(y) &= P(Y \leqq y) \\ &= P(|X| \leqq y) \\ &= P(-y \leqq X \leqq y) \\ &= F(y) - F(-y) \end{aligned}$$

y, después de la diferenciación,

$$g(y) = f(y) + f(-y)$$

También, puesto que $|x|$ no puede ser negativo, $g(y) = 0$ para $y < 0$. Arbitrariamente haciendo $g(0) = 0$, así podemos escribir

$$g(y) = \begin{cases} f(y) + f(-y) & \text{para } y > 0 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

Si X tiene la distribución normal estándar y $Y = |X|$, se sigue que

$$\begin{aligned} g(y) &= n(y; 0, 1) + n(-y; 0, 1) \\ &= 2n(y; 0, 1) \end{aligned}$$

para $y > 0$ y $g(y) = 0$ en cualquier otra parte. En el ejemplo 7.9 se puede encontrar una aplicación importante para este resultado. ▲

EJEMPLO 7.3

Si la densidad conjunta de X_1 y X_2 está dada por

$$f(x_1, x_2) = \begin{cases} 6e^{-3x_1 - 2x_2} & \text{para } x_1 > 0, x_2 > 0 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

encuentre la densidad de probabilidad de $Y = X_1 + X_2$.

Solución

Al integrar la densidad conjunta sobre la región sombreada de la figura 7.1, obtenemos

$$\begin{aligned} F(y) &= \int_0^y \int_0^{y-x_2} 6e^{-3x_1 - 2x_2}\, dx_1\, dx_2 \\ &= 1 + 2e^{-3y} - 3e^{-2y} \end{aligned}$$

y, al diferenciar con respecto a y, obtenemos

$$f(y) = 6(e^{-2y} - e^{-3y})$$

para $y > 0$; en cualquier otra parte, $f(y) = 0$. ▲

x_2

$x_1 + x_2 = y$

x_1

Figura 7.1 Diagrama para el ejemplo 7.3.

EJERCICIOS

7.1 Si la densidad de probabilidad de X está dada por

$$f(x) = \begin{cases} 2xe^{-x^2} & \text{para } x > 0 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

y $Y = X^2$, encuentre

(a) la función de distribución de Y;

(b) la densidad de probabilidad de Y.

7.2 Si X tiene una distribución exponencial con el parámetro θ, use la técnica de la función de distribución para encontrar la densidad de probabilidad de la variable aleatoria $Y = \ln X$.

7.3 Si X tiene la densidad uniforme con los parámetros $\alpha = 0$ y $\beta = 1$, use la técnica de la función de distribución para encontrar la densidad de probabilidad de la variable aleatoria $Y = \sqrt{X}$.

7.4 Si la densidad de probabilidad conjunta de X y Y está dada por

$$f(x, y) = \begin{cases} 4xye^{-(x^2+y^2)} & \text{para } x > 0, y > 0 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

y $Z = \sqrt{X^2 + Y^2}$, encuentre

(a) la función de distribución de Z;

(b) la densidad de probabilidad de Z.

7.5 Si X_1 y X_2 son variables aleatorias independientes que tienen densidades exponenciales con parámetros θ_1 y θ_2, use la técnica de la función de distribución para encontrar la densidad de probabilidad de $Y = X_1 + X_2$ cuando.

(a) $\theta_1 \neq \theta_2$;

(b) $\theta_1 = \theta_2$.

(El ejemplo 7.3 es un caso especial de esto con $\theta_1 = \frac{1}{3}$ y $\theta_2 = \frac{1}{2}$.)

7.6 Con respecto a las dos variables aleatorias del ejercicio 7.5, muestre que si $\theta_1 = \theta_2 = 1$, la variable aleatoria

$$Z = \frac{X_1}{X_1 + X_2}$$

tiene la densidad uniforme con $\alpha = 0$ y $\beta = 1$.

7.7 Sean X_1 y X_2 variables aleatorias independientes que tienen la densidad uniforme con $\alpha = 0$ y $\beta = 1$. Vaya a la figura 7.2, encuentre expresiones para la función de distribución de $Y = X_1 + X_2$ para

(a) $y \leqq 0$;

(b) $0 < y < 1$;

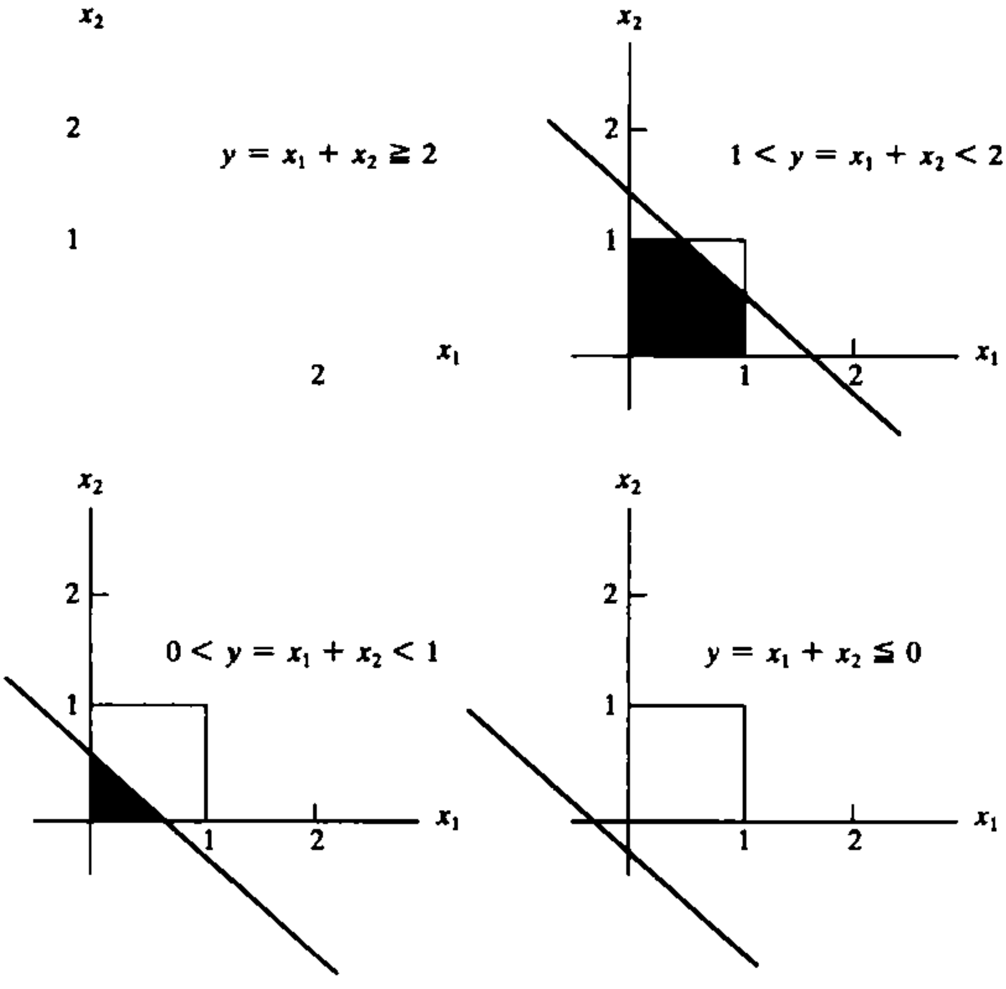

Figura 7.2 Diagrama para el ejercicio 7.7.

(c) $1 < y < 2$;

(d) $y \geqq 2$.

Encuentre también la densidad de probabilidad de Y.

7.8 Si la densidad conjunta de X y Y está dada por

$$f(x, y) = \begin{cases} e^{-(x+y)} & \text{para } x > 0, y > 0 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

y $Z = \dfrac{X + Y}{2}$, encuentre la densidad de probabilidad de Z por la técnica de la función de distribución.

APPLICACIONES

7.9 En el ejercicio 3.88, el precio de cierta mercancía (en dólares) y sus ventas totales (en unidades de 10,000) se denotó mediante P y S. Use la densidad conjunta dada en ese ejercicio y la técnica de la función de distribución para encontrar la densidad de probabilidad de $V = SP$, la cantidad total de dinero (en unidades de \$10,000) que se gasta en esa mercancía.

7.10 Con respecto al ejercicio 3.53, encuentre la densidad de probabilidad del millaje promedio de dos neumáticos como ésos. Suponga independencia.

7.11 En el ejercicio 3.107, X es la cantidad de dinero (en dólares) que un vendedor gasta en gasolina, y Y es la cantidad de dinero que se le reembolsa. Use la densidad de probabilidad conjunta dada en ese ejercicio y la técnica de la función de distribución para encontrar la densidad de probabilidad de la variable aleatoria $Z = X - Y$, la cantidad de dinero que no se le reembolsa.

7.12 Sea X la cantidad de gasolina premium (en 1,000 galones) que una estación de servicio tiene en sus tanques al principio del día, y Y la cantidad que esa estación de servicio vende durante el día. Si la densidad conjunta de X y Y está dada por

$$f(x, y) = \begin{cases} \dfrac{1}{200} & \text{para } 0 < y < x < 20 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

use la técnica de la función de distribución para encontrar la densidad de probabilidad de la cantidad que le queda a la estación de servicio en sus tanques al final del día.

7.13 Si los porcentajes de cobre y hierro en cierta clase de mineral son, respectivamente, X_1 y X_2. Si la densidad conjunta de estas dos variables aleatorias está dada por

$$f(x_1, x_2) = \begin{cases} \dfrac{3}{11}(5x_1 + x_2) & \text{para } x_1 > 0, x_2 > 0 \text{ y } x_1 + 2x_2 < 2 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

use la técnica de la función de distribución para encontrar la densidad de probabilidad de $Y = X_1 + X_2$. Encuentre también $E(Y)$, el porcentaje total esperado de cobre y hierro en el mineral.

7.3 TÉCNICA DE TRANSFORMACIÓN: UNA VARIABLE

Mostremos ahora cómo se puede determinar la densidad o distribución de probabilidad de la función de una variable aleatoria sin obtener primero su función de distribución. En el caso discreto no hay un problema real mientras la relación entre los valores de X y $Y = u(X)$ sea unívoca; todo lo que tenemos que hacer es la sustitución apropiada.

EJEMPLO 7.4

Si X es el número de caras obtenidas en cuatro tiros de un dado balanceado, encuentre la distribución de probabilidad de $Y = \dfrac{1}{1+X}$.

Solución

Usamos la fórmula para la distribución binomial con $n = 4$ y $\theta = \frac{1}{2}$, encontramos que la distribución de probabilidad de X está dada por

x	0	1	2	3	4
$f(x)$	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$

Entonces, usamos la relación $y = \dfrac{1}{1+x}$ para sustituir los valores de Y en vez de los valores de X, encontramos que la distribución de probabilidad de Y está dada por

y	1	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{4}$	$\frac{1}{5}$
$g(y)$	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$

Si hubiésemos querido hacer la sustitución directamente en la fórmula para la distribución binomial con $n = 4$ y $\theta = \frac{1}{2}$, podíamos haber sustituido $x = \dfrac{1}{y} - 1$ en vez de x en

$$f(x) = \binom{4}{x}\left(\frac{1}{2}\right)^4 \qquad \text{para } x = 0, 1, 2, 3, 4$$

y obtenemos

$$g(y) = f\left(\frac{1}{y} - 1\right) = \binom{4}{\frac{1}{y} - 1}\left(\frac{1}{2}\right)^4 \qquad \text{para } y = 1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5} \qquad \blacktriangle$$

Observe que en el ejemplo anterior las probabilidades quedaron sin cambio; la única diferencia es que en el resultado están relacionadas con los diversos valores de Y

en vez de los correspondientes valores de X. Esto es todo lo que hay para la **técnica de la transformación** (o **cambio de variable**) en el caso discreto mientras la relación sea unívoca. Si la relación no es unívoca, podemos proceder como en el siguiente ejemplo.

EJEMPLO 7.5

Con respecto al ejemplo 7.4, encuentre la distribución de probabilidad de la variable aleatoria $Z = (X - 2)^2$.

Solución

Calculamos las probabilidades $h(z)$ asociadas con los diversos valores de Z, y obtenemos

$$h(0) = f(2) = \frac{6}{16}$$

$$h(1) = f(1) + f(3) = \frac{4}{16} + \frac{4}{16} = \frac{8}{16}$$

$$h(4) = f(0) + f(4) = \frac{1}{16} + \frac{1}{16} = \frac{2}{16}$$

y por tanto

z	0	1	4
$h(z)$	$\frac{3}{8}$	$\frac{4}{8}$	$\frac{1}{8}$

▲

Para efectuar una transformación de la variable en el caso continuo, supondremos que la función dada por $y = u(x)$ es diferenciable ya sea creciente o decreciente para todos los valores dentro del intervalo de X para el cual $f(x) \neq 0$, de manera que la función inversa, dada por $x = w(y)$, existe para todos los valores correspondientes de y y es diferenciable excepto donde $u'(x) = 0$.† Bajo estas condiciones, podemos probar el siguiente teorema

TEOREMA 7.1 Sea $f(x)$ el valor de la densidad de probabilidad de la variable aleatoria continua X en x. Si la función dada por $y = u(x)$ es diferenciable y ya sea creciente o decreciente para todos los valores dentro del intervalo de X para el cual $f(x) \neq 0$, entonces, para estos valores de x, la ecuación $y = u(x)$ se puede despejar de manera única en x para $x = w(y)$, y para los valores correspondientes de y la densidad de probabilidad de $Y = u(X)$ está dada por

† Para evitar puntos donde $u'(x)$ podría ser 0, generalmente no incluimos los puntos terminales de los intervalos para los cuales las densidades de probabilidad no son cero. Ésta es la práctica que hemos seguido y continuaremos siguiendo en todo este libro.

$$g(y) = f[w(y)] \cdot |w'(y)| \qquad \text{siempre que } u'(x) \neq 0$$

En cualquier otra parte, $g(y) = 0$.

Demostración. Primero, demostremos el caso donde la función dada por $y = u(x)$ es creciente. Como se puede ver en la figura 7.3, X debe asumir un valor entre $w(a)$ y $w(b)$ cuando Y asume un valor entre a y b. Por tanto,

$$\begin{aligned} P(a < Y < b) &= P[w(a) < X < w(b)] \\ &= \int_{w(a)}^{w(b)} f(x)\,dx \\ &= \int_a^b f[w(y)]w'(y)\,dy \end{aligned}$$

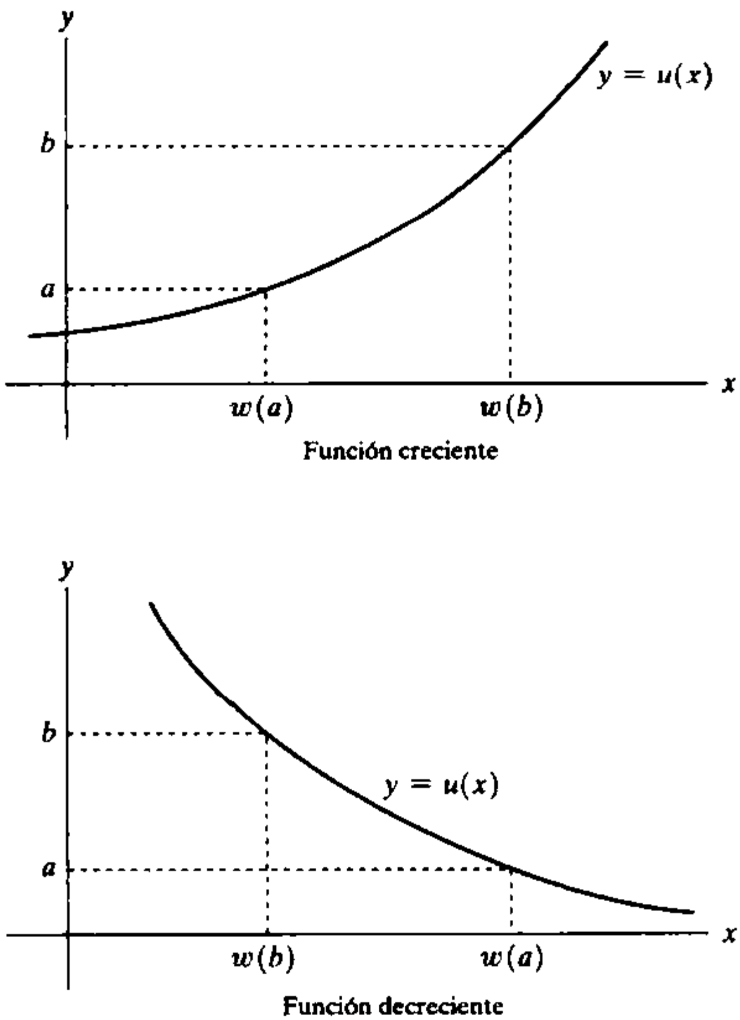

Figura 7.3 Diagramas para la prueba del teorema 7.1.

donde efectuamos el cambio de variable $y = u(x)$, o en forma equivalente $x = w(y)$, en la integral. De acuerdo con la definición 3.4, el integrando da la densidad de probabilidad de Y mientras $w'(y)$ exista, y podemos escribir

$$g(y) = f[w(y)]w'(y)$$

Cuando la función dada por $y = u(x)$ es decreciente, se puede ver de la figura 7.3 que X debe asumir un valor entre $w(b)$ y $w(a)$ cuando Y asume un valor entre a y b. Por tanto

$$\begin{aligned} P(a < Y < b) &= P[w(b) < X < w(a)] \\ &= \int_{w(b)}^{w(a)} f(x)\,dx \\ &= \int_{b}^{a} f[w(y)]w'(y)\,dy \\ &= -\int_{a}^{b} f[w(y)]w'(y)\,dy \end{aligned}$$

donde efectuamos el mismo cambio de variable como antes, y se sigue que

$$g(y) = -f[w(y)]w'(y)$$

Puesto que $w'(y) = \dfrac{dx}{dy} = \dfrac{1}{\dfrac{dy}{dx}}$ es positivo cuando la función dada por $y = u(x)$ es creciente, $-w'(y)$ es positivo cuando la función dada por $y = u(x)$ es decreciente, podemos combinar los dos casos al escribir

$$g(y) = f[w(y)] \cdot |w'(y)| \qquad \blacktriangledown$$

EJEMPLO 7.6

Si X tiene la distribución exponencial dada por

$$f(x) = \begin{cases} e^{-x} & \text{para } x > 0 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

encuentre la densidad de probabilidad de la variable aleatoria $Y = \sqrt{X}$.

Solución

La ecuación $y = \sqrt{x}$, que relaciona los valores de X y Y, tiene la inversa única $x = y^2$, que da $w'(y) = \dfrac{dx}{dy} = 2y$. Por consiguiente,

$$g(y) = e^{-y^2}|2y| = 2ye^{-y^2}$$

para $y > 0$ de acuerdo al teorema 7.1. Puesto que la probabilidad de obtener un valor de Y menor que o igual a 0, como la probabilidad de obtener un valor de X menor que o igual a 0, es cero, se sigue que la densidad de probabilidad de Y está dada por

$$g(y) = \begin{cases} 2ye^{-y^2} & \text{para } y > 0 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

Advierta que ésta es la distribución de Weibull del ejercicio 6.23 con $\alpha = 1$ y $\beta = 2$. ▲

Los dos diagramas de la figura 7.4 ilustran lo que pasó en este ejemplo cuando transformamos de X a Y. Como en el caso discreto (por ejemplo, el ejemplo 7.4), las probabilidades permanecen igual, pero corresponden a diferentes valores (intervalos de valores) de las variables aleatorias respectivas. En el diagrama de la izquierda, la probabilidad 0.35 corresponde al evento de que X asumirá un valor en el intervalo 1 a 4, y en el diagrama de la derecha, la probabilidad 0.35 corresponde al evento de que Y asumirá un valor en el intervalo de 1 a 2.

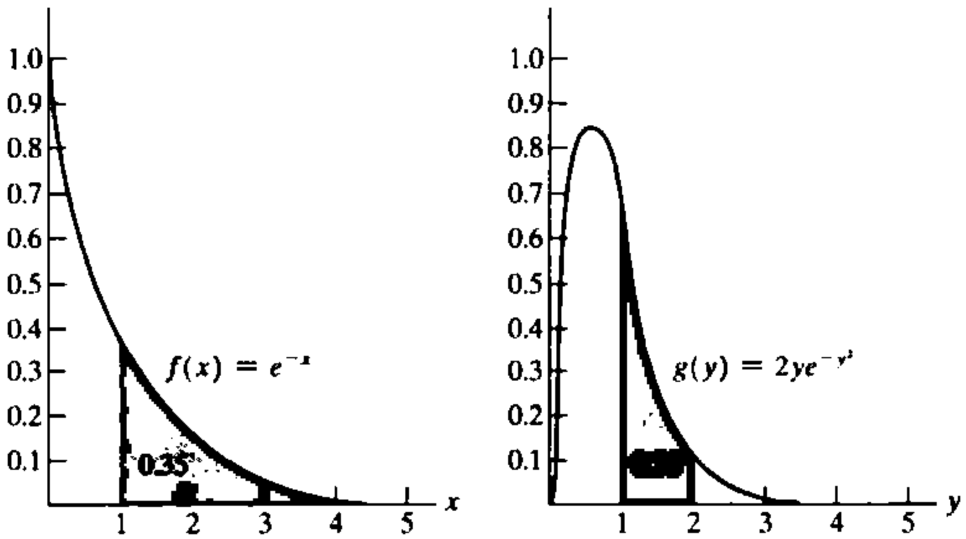

Figura 7.4 Diagramas para el ejemplo 7.6.

EJEMPLO 7.7

Si se hace girar la flecha doble de la figura 7.5 de manera que la variable aleatoria Θ tenga la densidad uniforme

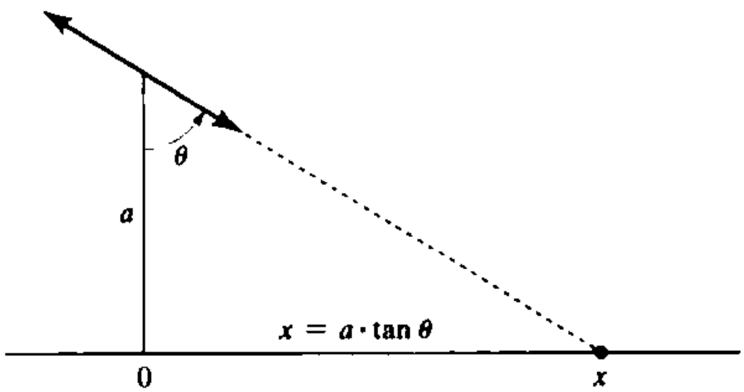

Figura 7.5 Diagrama para el ejemplo 7.7.

$$f(\theta) = \begin{cases} \dfrac{1}{\pi} & \text{para } -\dfrac{\pi}{2} < \theta < \dfrac{\pi}{2} \\ 0 & \text{en cualquier otra parte} \end{cases}$$

determine la densidad de probabilidad de X, la abscisa del punto en el eje x a la cual apuntará la flecha.

Solución

Como es evidente a partir del diagrama, la relación entre x y θ está dada por $x = a \cdot \tan \theta$, de manera que

$$\frac{d\theta}{dx} = \frac{a}{a^2 + x^2}$$

y se sigue que

$$\begin{aligned} g(x) &= \frac{1}{\pi} \cdot \left| \frac{a}{a^2 + x^2} \right| \\ &= \frac{1}{\pi} \cdot \frac{a}{a^2 + x^2} \qquad \text{para } -\infty < x < \infty \end{aligned}$$

de acuerdo al teorema 7.1. Observe que éste es un caso especial de la distribución de Cauchy del ejercicio 6.6. ▲

EJEMPLO 7.8

Si $F(x)$ es el valor de la función de distribución de la variable aleatoria continua X en x, encuentre la densidad de probabilidad de $Y = F(X)$.

Solución

Como se puede ver a partir de la figura 7.6, el valor de Y que corresponde a cualquier valor particular de X está dado por el área bajo la curva, esto es, el área ba-

x

Figura 7.6 Diagrama para el ejemplo 7.8.

jo la gráfica de la densidad de X a la izquierda de x. Diferenciamos $y = F(x)$ con respecto a x, y obtenemos

$$\frac{dy}{dx} = F'(x) = f(x)$$

y por tanto

$$\frac{dx}{dy} = \frac{1}{\frac{dy}{dx}} = \frac{1}{f(x)}$$

siempre que $f(x) \neq 0$. Se sigue por el teorema 7.1 que

$$g(y) = f(x) \cdot \left| \frac{1}{f(x)} \right| = 1$$

para $0 < y < 1$, y podemos decir que y tiene la densidad uniforme con $\alpha = 0$ y $\beta = 1$. ▲

La transformación que efectuamos en este ejemplo se llama **transformación de la integral de probabilidad**. El resultado no es sólo de importancia teórica, sino que facilita la **simulación** de los valores observados de variables aleatorias continuas. En la página 265 se da una referencia de cómo se hace, especialmente en relación con la distribución normal.

Cuando no se satisfacen las condiciones que sustentan el teorema 7.1, podemos estar en dificultades serias y tendríamos que usar el método de la sección 7.2 o una generalización del teorema 7.1 al que nos referimos entre las referencias de la página 265; algunas veces hay una salida fácil, como en el siguiente ejemplo.

EJEMPLO 7.9

Si X tiene una distribución normal estándar, encuentre la densidad de probabilidad de $Z = X^2$.

Solución

Puesto que la función dada por $z = x^2$ es decreciente para valores negativos de x, y creciente para los valores positivos de x, no se satisfacen las condiciones del teorema 7.1. Sin embargo, es posible hacer en dos pasos la transformación de X a Z: primero, encontramos la densidad de probabilidad de $Y = |X|$, y después encontramos la densidad de probabilidad de $Z = Y^2$ $(= X^2)$.

En lo tocante al primer paso, ya hemos estudiado la transformación $Y = |X|$ en el ejemplo 7.2; de hecho, demostramos que si X tiene la distribución normal estándar, entonces $Y = |X|$ tiene la densidad de probabilidad

$$g(y) = 2n(y; 0, 1) = \frac{2}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}$$

para $y > 0$, y $g(y) = 0$ en cualquier otra parte. Para el segundo paso, la función dada por $z = y^2$ es creciente para $y > 0$, esto es, para todos los valores de Y para lo cuales $g(y) \neq 0$. Así, podemos usar el teorema 7.1, y puesto que

$$\frac{dy}{dz} = \frac{1}{2} z^{-\frac{1}{2}}$$

obtenemos

$$h(z) = \frac{2}{\sqrt{2\pi}} e^{-\frac{1}{2}z} \left| \frac{1}{2} z^{-\frac{1}{2}} \right|$$

$$= \frac{1}{\sqrt{2\pi}} z^{-\frac{1}{2}} e^{-\frac{1}{2}z}$$

para $z > 0$, y $h(z) = 0$ en cualquier otra parte. Observe que puesto que $\Gamma(\frac{1}{2}) = \sqrt{\pi}$, la distribución a la que hemos llegado para Z es una distribución ji cuadrada (véase la definición 6.4) con $\nu = 1$. ▲

7.4 TÉCNICA DE TRANSFORMACIÓN: VARIAS VARIABLES

El método de la sección precedente también se puede usar para encontrar la distribución de una variable aleatoria que es una función de dos o más variables aleatorias. Supongamos, por ejemplo, que se nos da la distribución conjunta de dos variables aleatorias X_1 y X_2 y que queremos determinar la distribución de probabilidad o la densidad de probabilidad de la variable aleatoria $Y = u(X_1, X_2)$. Si la relación entre y y x_1 con x_2 mantenida constante o la relación entre y y x_2 con x_1 mantenida constante lo permite, podemos proceder en el caso discreto como en el ejemplo 7.4 para encontrar la distribución conjunta de Y y X_2 o la de X_1 y Y y luego sumarles los valores de la otra variable aleatoria para obtener la distribución marginal de Y. En el caso continuo, primero usamos el teorema 7.1 con la fórmula de transformación escrita como

$$g(y, x_2) = f(x_1, x_2) \cdot \left| \frac{\partial x_1}{\partial y} \right|$$

o como

$$g(x_1, y) = f(x_1, x_2) \cdot \left| \frac{\partial x_2}{\partial y} \right|$$

donde $f(x_1, x_2)$ y la derivada parcial se deben expresar en términos de y y x_2 o x_1 y y. entonces sacamos por integración la otra variable para obtener la densidad marginal de y.

EJEMPLO 7.10

Si X_1 y X_2 son variables aleatorias independientes que tienen distribuciones de Poisson con los parámetros λ_1 y λ_2, encuentre la distribución de probabilidad de la variable aleatoria $Y = X_1 + X_2$.

Solución

Puesto que X_1 y X_2 son independientes, su distribución conjunta está dada por

$$f(x_1, x_2) = \frac{e^{-\lambda_1}(\lambda_1)^{x_1}}{x_1!} \cdot \frac{e^{-\lambda_2}(\lambda_2)^{x_2}}{x_2!}$$
$$= \frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1)^{x_1}(\lambda_2)^{x_2}}{x_1!x_2!}$$

para $x_1 = 0, 1, 2, \ldots$ y $x_2 = 0, 1, 2, \ldots$. Puesto que $y = x_1 + x_2$ y por tanto $x_1 = y - x_2$, podemos sustituir $y - x_2$ en vez de x_1, obteniendo

$$g(y, x_2) = \frac{e^{-(\lambda_1+\lambda_2)}(\lambda_2)^{x_2}(\lambda_1)^{y-x_2}}{x_2!(y - x_2)!}$$

para $y = 0, 1, 2, \ldots$ y $x_2 = 0, 1, \ldots, y$, para la distribución conjunta de Y y X_2. Entonces, al sumar sobre x_2 de 0 a y, obtenemos

$$h(y) = \sum_{x_2=0}^{y} \frac{e^{-(\lambda_1+\lambda_2)}(\lambda_2)^{x_2}(\lambda_1)^{y-x_2}}{x_2!(y - x_2)!}$$
$$= \frac{e^{-(\lambda_1+\lambda_2)}}{y!} \cdot \sum_{x_2=0}^{y} \frac{y!}{x_2!(y - x_2)!}(\lambda_2)^{x_2}(\lambda_1)^{y-x_2}$$

después sacamos como factor $e^{-(\lambda_1+\lambda_2)}$ y multiplicamos y dividimos por $y!$. Identificamos la suma a la que hemos llegado como la expansión binomial de $(\lambda_1 + \lambda_2)^y$, y finalmente obtenemos

$$h(y) = \frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1 + \lambda_2)^y}{y!} \qquad \text{para } y = 0, 1, 2, \ldots$$

y hemos mostrado así que la suma de dos variables aleatorias independientes que tienen distribuciones de Poisson con los parámetros λ_1 y λ_2 tiene una distribución de Poisson con el parámetro $\lambda = \lambda_1 + \lambda_2$. ▲

EJEMPLO 7.11

Si la densidad de probabilidad conjunta de X_1 y X_2 está dada por

$$f(x_1, x_2) = \begin{cases} e^{-(x_1+x_2)} & \text{para } x_1 > 0, x_2 > 0 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

encuentre la densidad de probabilidad de $Y = \dfrac{X_1}{X_1 + X_2}$.

Solución

Puesto que y decrece cuando x_2 aumenta y x_1 se mantiene constante, podemos usar el teorema 7.1 (como se modificó en la página 249) para encontrar la densidad conjunta de X_1 y Y. Puesto que $y = \dfrac{x_1}{x_1 + x_2}$ nos da $x_2 = x_1 \cdot \dfrac{1 - y}{y}$ y por tanto

$$\frac{\partial x_2}{\partial y} = -\frac{x_1}{y^2}$$

se sigue que

$$g(x_1, y) = e^{-x_1/y}\left|-\frac{x_1}{y^2}\right| = \frac{x_1}{y^2}\cdot e^{-x_1/y}$$

para $x_1 > 0$ y $0 < y < 1$. Finalmente, al sacar x_1 por integración y cambiar la variable de integración a $u = x_1/y$, obtenemos

$$\begin{aligned} h(y) &= \int_0^\infty \frac{x_1}{y^2}\cdot e^{-x_1/y}\,dx_1 \\ &= \int_0^\infty u\cdot e^{-u}\,du \\ &= \Gamma(2) \\ &= 1 \end{aligned}$$

para $0 < y < 1$ y $h(y) = 0$ en cualquier otra parte. Así, la variable aleatoria Y tiene la densidad uniforme con $\alpha = 0$ y $\beta = 1$. (Advierta que en el ejercicio 7.6 se le pidió que mostrara esto con la técnica de la función de distribución.) ▲

El ejemplo anterior también se podía haber trabajado con un método general donde empezamos con la distribución conjunta de dos variables aleatorias X_1 y X_2, y determinamos la distribución conjunta de dos variables aleatorias nuevas $Y_1 = u_1(X_1, X_2)$ y $Y_2 = u_2(X_1, X_2)$. Entonces podemos encontrar la distribución marginal de Y_1 y Y_2 por suma o integración.

Este método se usa principalmente en el caso continuo, donde necesitamos el siguiente teorema, que es una generalización directa del teorema 7.1.

TEOREMA 7.2 Sea $f(x_1, x_2)$ el valor de la densidad de probabilidad conjunta de las variables aleatorias continuas X_1 y X_2 en (x_1, x_2). Si las funciones dadas por $y_1 = u_1(x_1, x_2)$ y $y_2 = u_2(x_1, x_2)$ son parcialmente diferenciables con respecto tanto a x_1 como a x_2 representan una transformación unívoca para todos los valores dentro del intervalo de X_1 y X_2 para los cuales $f(x_1, x_2) \neq 0$, entonces, para estos valores de x_1 y x_2, las ecuaciones $y_1 = u_1(x_1, x_2)$ y $y_2 = u_2(x_1, x_2)$ se pueden resolver de manera única para x_1 y x_2 a fin de dar $x_1 = w_1(y_1, y_2)$ y $x_2 = w_2(y_1, y_2)$, y para los valores correspondientes de y_1 y y_2, la densidad de probabilidad conjunta de $Y_1 = u_1(X_1, X_2)$ y $Y_2 = u_2(X_1, X_2)$ está dada por

$$g(y_1, y_2) = f[w_1(y_1, y_2), w_2(y_1, y_2)]\cdot|J|$$

Aquí, J, llamado el **jacobiano** de la transformación, es el determinante

$$J = \begin{vmatrix} \dfrac{\partial x_1}{\partial y_1} & \dfrac{\partial x_1}{\partial y_2} \\ \dfrac{\partial x_2}{\partial y_1} & \dfrac{\partial x_2}{\partial y_2} \end{vmatrix}$$

En cualquier otra parte, $g(y_1, y_2) = 0$.

No demostraremos este teorema, pero la información sobre los jacobianos y su aplicación se puede encontrar en la mayoría de los libros de texto de cálculo avanzado. Se usan principalmente en relación con integrales múltiples, digamos, cuando queremos cambiar de coordenadas rectangulares a coordenadas polares o de coordenadas rectangulares a coordenadas esféricas.

EJEMPLO 7.12

Con respecto a las variables aleatorias X_1 y X_2 del ejemplo 7.11, encuentre

(a) la densidad conjunta de $Y_1 = X_1 + X_2$ y $Y_2 = \dfrac{X_1}{X_1 + X_2}$;

(b) la densidad marginal de Y_2.

Solución

(a) Despejamos $y_1 = x_1 + x_2$ y $y_2 = \dfrac{x_1}{x_1 + x_2}$ para x_1 y x_2, y obtenemos $x_1 = y_1 y_2$ y $x_2 = y_1(1 - y_2)$, y se sigue que

$$J = \begin{vmatrix} y_2 & y_1 \\ 1 - y_2 & -y_1 \end{vmatrix} = -y_1$$

Puesto que la transformación es unívoca, al representar la región $x_1 > 0$ y $x_2 > 0$ en el plano x_1x_2-en la región $y_1 > 0$ y $0 < y_2 < 1$ en el plano y_1y_2, podemos usar el teorema 7.2 y se sigue que

$$g(y_1, y_2) = e^{-y_1}|-y_1| = y_1 e^{-y_1}$$

para $y_1 > 0$ y $0 < y_2 < 1$; en cualquier otra parte $g(y_1, y_2) = 0$.

(b) Al usar la densidad conjunta obtenida en el inciso (a) y sacar y_1, por integración, obtenemos

$$\begin{aligned} h(y_2) &= \int_0^\infty g(y_1, y_2)\, dy_1 \\ &= \int_0^\infty y_1 e^{-y_1}\, dy_1 \\ &= \Gamma(2) \\ &= 1 \end{aligned}$$

para $0 < y_2 < 1$; en cualquier otra parte, $h(y_2) = 0$. Advierta que este resultado concuerda con el obtenido en la página 251. ▲

EJEMPLO 7.13

Si la densidad conjunta de X_1 y X_2 está dada por

$$f(x_1, x_2) = \begin{cases} 1 & \text{para } 0 < x_1 < 1, 0 < x_2 < 1 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

encuentre

(a) la densidad conjunta de $Y = X_1 + X_2$ y $Z = X_2$;
(b) la densidad marginal de Y.

Observe que en el ejercicio 7.7 se le pidió que trabajara el mismo problema mediante la técnica de la función de distribución.

Solución

(a) Al despejar $y = x_1 + x_2$ y $z = x_2$ para x_1 y x_2, obtenemos $x_1 = y - z$ y $x_2 = z$, de manera que

$$J = \begin{vmatrix} 1 & -1 \\ 0 & 1 \end{vmatrix} = 1$$

Debido a que esta transformación es unívoca, al representar la región $0 < x_1 < 1$ y $0 < x_2 < 1$ en el plano x_1x_2 en la región $z < y < z + 1$ y $0 < z < 1$ en el plano yz (véase la figura 7.7), podemos usar el teorema 7.2 y obtenemos

$$g(y, z) = 1 \cdot |1| = 1$$

para $z < y < z + 1$ y $0 < z < 1$; en cualquier otra parte $g(y, z) = 0$.

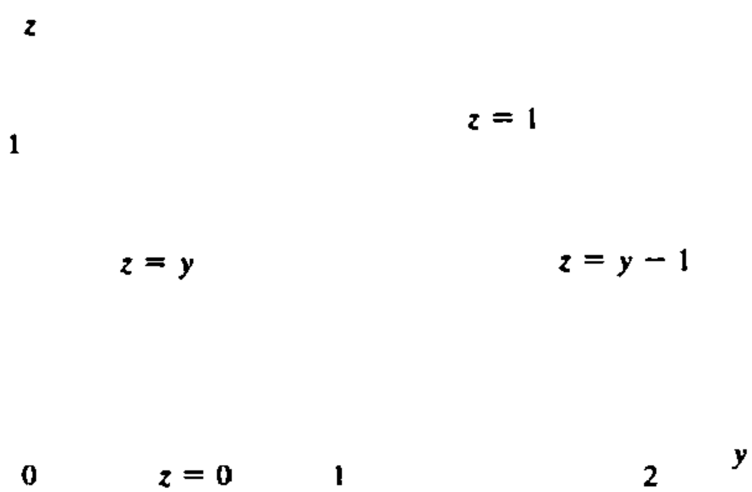

Figura 7.7 Espacio muestral transformado para el ejemplo 7.13.

(b) Al sacar z por integración de manera separada para $y \leqq 0$, $0 < y < 1$, $1 < y < 2$, y $y \geqq 2$, obtenemos

$$h(y) = \begin{cases} 0 & \text{para } y \leqq 0 \\ \displaystyle\int_0^y 1\,dz = y & \text{para } 0 < y < 1 \\ \displaystyle\int_{y-1}^1 1\,dz = 2 - y & \text{para } 1 < y < 2 \\ 0 & \text{para } y \geqq 2 \end{cases}$$

y para hacer continua la función de densidad, hacemos $h(1) = 1$. Así hemos demostrado que la suma de las variables aleatorias dadas tiene la **densidad de probabilidad triangular**, cuya gráfica se muestra en la figura 7.8. ▲

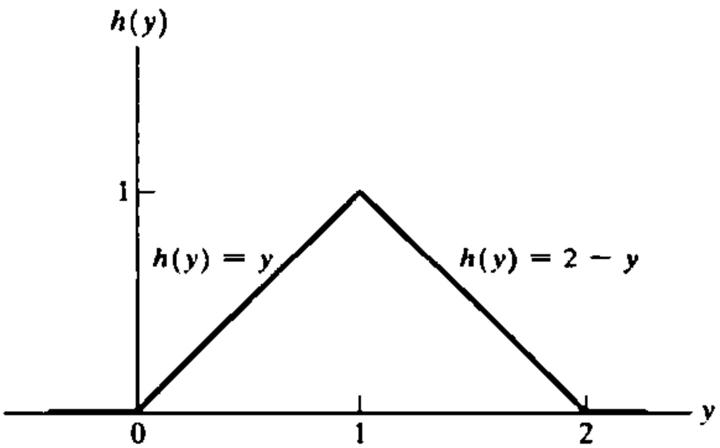

Figura 7.8 Densidad de probabilidad triangular.

Hasta ahora aquí sólo hemos considerado funciones de dos variables aleatorias, pero el método basado en el teorema 7.2 se puede generalizar fácilmente a funciones de tres o más variables aleatorias. Por ejemplo, si nos dan la densidad de probabilidad conjunta de tres variables aleatorias X_1, X_2 y X_3, y queremos encontrar la densidad de probabilidad conjunta de las variables aleatorias $Y_1 = u_1(X_1, X_2, X_3)$, $Y_2 = u_2(X_1, X_2, X_3)$ y $Y_3 = u_3(X_1, X_2, X_3)$, el enfoque general es el mismo, pero el jacobiano es ahora el determinante de 3×3

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \frac{\partial x_1}{\partial y_3} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \frac{\partial x_2}{\partial y_3} \\ \frac{\partial x_3}{\partial y_1} & \frac{\partial x_3}{\partial y_2} & \frac{\partial x_3}{\partial y_3} \end{vmatrix}$$

Una vez determinada la densidad de probabilidad conjunta de las tres nuevas variables aleatorias, podemos encontrar la densidad marginal de dos variables aleatorias cualquiera, o de una cualquiera, por integración.

EJEMPLO 7.14

Si la densidad de probabilidad conjunta de X_1, X_2 y X_3 está dada por

$$f(x_1, x_2, x_3) = \begin{cases} e^{-(x_1+x_2+x_3)} & \text{para } x_1 > 0, x_2 > 0, x_3 > 0 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

encuentre

(a) la densidad conjunta de $Y_1 = X_1 + X_2 + X_3$, $Y_2 = X_2$ y $Y_3 = X_3$;

(b) la densidad marginal de Y_1.

Solución

(a) Al resolver el sistema de ecuaciones $y_1 = x_1 + x_2 + x_3$, $y_2 = x_2$ y $y_3 = x_3$ para x_1, x_2 y x_3, obtenemos $x_1 = y_1 - y_2 - y_3$, $x_2 = y_2$ y $x_3 = y_3$. Se sigue que

$$J = \begin{vmatrix} 1 & -1 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = 1$$

y, puesto que la transformación es unívoca, que

$$\begin{aligned} g(y_1, y_2, y_3) &= e^{-y_1} \cdot |1| \\ &= e^{-y_1} \end{aligned}$$

para $y_2 > 0$, $y_3 > 0$ y $y_1 > y_2 + y_3$; en cualquier otra parte, $g(y_1, y_2, y_3) = 0$.

(b) Eliminamos y_2 y y_3, por integración, y obtenemos

$$\begin{aligned} h(y_1) &= \int_0^{y_1} \int_0^{y_1 - y_3} e^{-y_1}\, dy_2\, dy_3 \\ &= \frac{1}{2} y_1^2 \cdot e^{-y_1} \end{aligned}$$

para $y_1 > 0$; $h(y_1) = 0$ en cualquier otra parte. Observe que hemos mostrado que la suma de tres variables aleatorias independientes que tienen la distribución gamma con $\alpha = 1$ y $\beta = 1$ es una variable aleatoria que tiene la distribución gamma con $\alpha = 3$ y $\beta = 1$. ▲

Como el lector encontrará en el ejercicio 7.47, hubiera sido más fácil obtener el resultado del inciso (b) del ejemplo 7.14 usando el método basado en el teorema 7.1 como se modificó en la página 249.

EJERCICIOS

7.14 Si X tiene una distribución hipergeométrica con $k = 3$, $N = 6$ y $n = 2$, encuentre la distribución de probabilidad de Y, el número de éxitos menos el número de fracasos.

7.15 Con respecto al ejercicio 7.14, encuentre la distribución de probabilidad de la variable aleatoria $Z = (X - 1)^2$.

7.16 Si X tiene una distribución binomial con $n = 3$ y $\theta = \frac{1}{3}$, encuentre las distribuciones de probabilidad de

(a) $Y = \dfrac{X}{1 + X}$;

(b) $U = (X - 1)^4$.

7.17 Si X tiene una distribución geométrica con $\theta = \frac{1}{3}$, encuentre la fórmula para la distribución de probabilidad de la variable aleatoria $Y = 4 - 5X$.

7.18 Si X es el total que tiramos con un par de dados, para el cual la distribución de probabilidad se da en la página 77, encuentre la distribución de probabilidad del residuo que obtenemos cuando los valores de X se dividen entre 3.

7.19 Use la técnica de la transformación de la variable para probar el teorema 6.7.

7.20 Use la técnica de la transformación para rehacer el ejemplo 7.1.

7.21 Si $X = \ln Y$ tiene una distribución normal con la media μ ay la desviación estándar σ, encuentre la densidad de probabilidad de Y, la cual se dice que tiene **la distribución log-normal**.

7.22 Si $F(x)$ es el valor de la función de distribución de X en x, entonces la **mediana** de X, denotada por $\tilde{\mu}$ es tal que $F(\tilde{\mu}) = \frac{1}{2}$. Con respecto al ejercicio 7.21, muestre que $\tilde{\mu} = e^{\mu}$.

7.23 Con respecto al ejercicio 7.21, muestre que la distribución log-normal tiene un máximo relativo en $e^{\mu - \sigma^2}$.

7.24 Si la densidad de probabilidad de X está dada por

$$f(x) = \begin{cases} \dfrac{x}{2} & \text{para } 0 < x < 2 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

encuentre la densidad de probabilidad de $Y = X^3$. También, dibuje las gráficas de las densidades de probabilidad de X y Y e indique las áreas respectivas bajo las curvas que representan $P(\frac{1}{2} < X < 1)$ y $P(\frac{1}{8} < Y < 1)$.

7.25 Si la densidad de probabilidad de X está dada por

$$f(x) = \begin{cases} \dfrac{kx^3}{(1 + 2x)^6} & \text{para } x > 0 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

donde k es una constante apropiada, encuentre la densidad de probabilidad de la variable aleatoria $Y = \dfrac{2X}{1 + 2X}$. Identifique la distribución de Y, y determine así el valor de k.

7.26 Si X tiene la densidad uniforme con $\alpha = 0$ y $\beta = 1$, muestre que la variable aleatoria $Y = -2 \cdot \ln X$ tiene una distribución gamma. ¿Cuáles son sus parámetros?

7.27 Si X tiene la densidad uniforme con $\alpha = 0$ y $\beta = 1$, muestre que $Y = X^{-1/\alpha}$ con $\alpha > 0$ tiene la distribución de Pareto del ejercicio 6.21.

7.28 Considere la variable aleatoria X con la densidad de probabilidad de

$$f(x) = \begin{cases} \dfrac{3x^2}{2} & \text{para } -1 < x < 1 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

(a) Use el resultado del ejemplo 7.2 para encontrar la densidad de probabilidad de $Y = |X|$.

(b) Encuentre la densidad de probabilidad de $Z = X^2$ $(= Y^2)$.

7.29 Considere la variable aleatoria X con densidad uniforme que tiene $\alpha = 0$ y $\beta = 1$.

(a) Use el resultado del ejemplo 7.2 para encontrar la densidad de probabilidad de $Y = |X|$.

(b) Encuentre la densidad de probabilidad de $Z = X^4$ $(= Y^4)$.

7.30 Si la distribución de probabilidad conjunta de X_1 y X_2 está dada por

$$f(x_1, x_2) = \frac{x_1 x_2}{36}$$

para $x_1 = 1, 2, 3$ y $x_2 = 1, 2, 3$, encuentre

(a) la distribución de probabilidad de $X_1 X_2$;

(b) la distribución de probabilidad de X_1/X_2.

7.31 Con respecto al ejercicio 7.30, encuentre

(a) la probabilidad conjunta de $Y_1 = X_1 + X_2$ y $Y_2 = X_1 - X_2$;

(b) la distribución marginal de Y_1.

7.32 Si la distribución de probabilidad conjunta de X y Y está dada por

$$f(x, y) = \frac{(x - y)^2}{7}$$

para $x = 1, 2$ y $y = 1, 2, 3$, encuentre

(a) la distribución conjunta de $U = X + Y$ y $V = X - Y$;

(b) la distribución marginal de U.

7.33 Si X_1, X_2 y X_3 tiene la distribución multinomial (véase la definición 5.6) con $n = 2$, $\theta_1 = \frac{1}{4}$, $\theta_2 = \frac{1}{3}$ y $\theta_3 = \frac{5}{12}$, encuentre la distribución de probabilidad conjunta de $Y_1 = X_1 + X_2$, $Y_2 = X_1 - X_2$ y $Y_3 = X_3$.

7.34 Con respecto al ejercicio 3.12, encuentre

(a) la distribución de probabilidad de $U = X + Y$;

(b) la distribución de probabilidad de $V = XY$;

(c) la distribución de probabilidad de $W = X - Y$.

7.35 Si X_1 y X_2 son variables aleatorias independientes que tienen la distribución binomial con los respectivos parámetros n_1 y θ y n_2 y θ, muestre que $Y = X_1 + X_2$ tiene la distribución binomial con los parámetros $n_1 + n_2$ y θ. (*Sugerencia*: use el teorema 1.12.)

7.36 Si X_1 y X_2 son variables aleatorias independientes que tienen la distribución geométrica con el parámetro θ, muestre que $Y = X_1 + X_2$ es una variable aleatoria que tiene la distribución binomial negativa con los parámetros θ y $k = 2$.

7.37 Si X_1 y Y son variables aleatorias independientes que tienen la distribución normal estándar, muestre que la variable aleatoria $Z = X + Y$ también está distribuida normalmente. (*Sugerencia*: complete el cuadrado en el exponente.) ¿Cuáles son la media y la varianza de esta distribución normal?

7.38 Considere dos variables aleatorias X y Y con la densidad de probabilidad conjunta

$$f(x, y) = \begin{cases} 12xy(1 - y) & \text{para } 0 < x < 1, 0 < y < 1 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

Encuentre la densidad de probabilidad de $Z = XY^2$ mediante el teorema 7.1 (con la modificación de la página 249), a fin de determinar la densidad de probabilidad conjunta de Y y Z y después integrar para eliminar y.

7.39 Rehaga el ejercicio 7.38 mediante el uso del teorema 7.2 para determinar la densidad de probabilidad conjunta de $Z = XY^2$ y $U = Y$ y después encuentre la densidad marginal de Z.

7.40 Considere dos variables aleatorias independientes X_1 y X_2 que tienen la misma distribución de Cauchy

$$f(x) = \frac{1}{\pi(1 + x^2)} \qquad \text{para } -\infty < x < \infty$$

Encuentre la densidad de probabilidad de $Y_1 = X_1 + X_2$ mediante el teorema 7.1 (como se modificó en la página 249) para determinar la densidad de probabilidad conjunta de X_1 y Y_1 y después integrar para eliminar x_1. También, identifique la distribución de Y_1.

7.41 Rehaga el ejercicio 7.40 mediante el uso del teorema 7.2 para determinar la densidad de probabilidad conjunta de $Y_1 = X_1 + X_2$ y $Y_2 = X_1 - X_2$ y después encuentre la densidad marginal de Y_1.

7.42 Considere dos variables aleatorias X y Y cuya densidad de probabilidad conjunta está dada por

$$f(x, y) = \begin{cases} \frac{1}{2} & \text{para } x > 0, y > 0, x + y < 2 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

Encuentre la densidad de probabilidad de $U = Y - X$ al usar el teorema 7.1 como se modificó en la página 249.

7.43 Rehaga el ejercicio 7.42 mediante el teorema 7.2 para determinar la densidad de probabilidad conjunta de $U = Y - X$ y $V = X$ y después encuentre la densidad marginal de U.

7.44 Sean X_1 y X_2 dos variables aleatorias continuas que tienen la densidad de probabilidad conjunta

$$f(x_1, x_2) = \begin{cases} 4x_1x_2 & \text{para } 0 < x_1 < 1, 0 < x_2 < 1 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

Encuentre la densidad de probabilidad conjunta de $Y_1 = X_1^2$ y $Y_2 = X_1X_2$.

7.45 Sean X y Y dos variables aleatorias continuas que tienen la densidad de probabilidad conjunta

$$f(x, y) = \begin{cases} 24xy & \text{para } 0 < x < 1, 0 < y < 1, x + y < 1 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

Encuentre la densidad de probabilidad conjunta de $Z = X + Y$ y $W = X$.

7.46 Sean X y Y dos variables aleatorias independientes que tienen distribuciones gamma idénticas

(a) Encuentre la densidad de probabilidad conjunta de las variables aleatorias $U = \dfrac{X}{X + Y}$ y $V = X + Y$.

(b) Encuentre e identifique la densidad marginal de U.

7.47 En la página 249 indicamos que el método de la transformación basado en el teorema 7.1 se puede generalizar para que también se aplique a variables aleatorias que son funciones de dos o más variables aleatorias. Hasta ahora hemos usado este método sólo para funciones de dos variables aleatorias, pero cuando hay tres, por ejemplo, introducimos la nueva variable aleatoria en lugar de una de las variables aleatorias originales, y después eliminamos (por suma o integración) las otras dos variables aleatorias con que empezamos. Use este método para rehacer el ejemplo 7.14.

7.48 En el ejemplo 7.13 encontramos la densidad de probabilidad de la suma de dos variables aleatorias independientes que tienen la densidad uniforme con $\alpha = 0$ y $\beta = 1$. Dada una tercera variable aleatoria X_3, la cual tiene la misma densidad uniforme y es independiente tanto de X_1 como de X_2, demuestre que si $U = Y + X_3 = X_1 + X_2 + X_3$, entonces

(a) la densidad de probabilidad conjunta de U y Y está dada por

$$g(u, y) = \begin{cases} y & \text{para las regiones I y II de la figura 7.9} \\ 2 - y & \text{para las regiones III y IV de la figura 7.9} \\ 0 & \text{en cualquier otra parte} \end{cases}$$

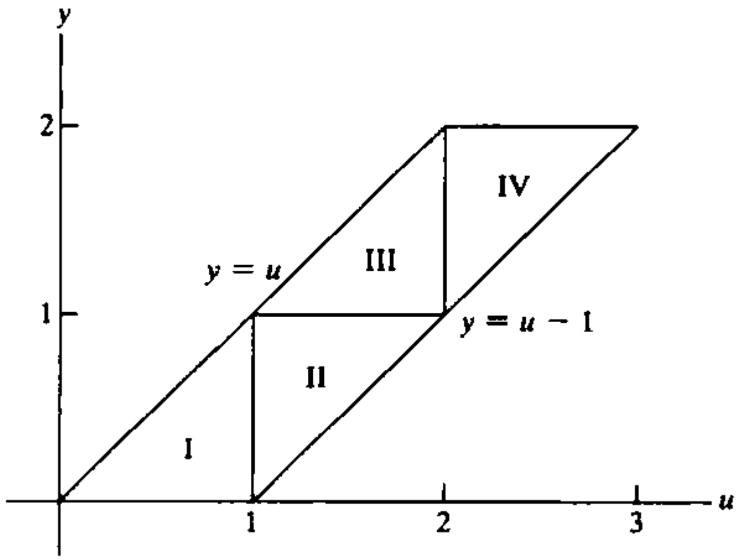

Figura 7.9 Diagrama para el ejercicio 7.48.

(b) la densidad de probabilidad de U está dada por

$$h(u) = \begin{cases} 0 & \text{para } u \leqq 0 \\ \dfrac{1}{2}u^2 & \text{para } 0 < u < 1 \\ \dfrac{1}{2}u^2 - \dfrac{3}{2}(u-1)^2 & \text{para } 1 < u < 2 \\ \dfrac{1}{2}u^2 - \dfrac{3}{2}(u-1)^2 + \dfrac{3}{2}(u-2)^2 & \text{para } 2 < u < 3 \\ 0 & \text{para } u \geqq 3 \end{cases}$$

Advierta que si $h(1) = h(2) = \frac{1}{2}$, esto hará continua la densidad de probabilidad de U.

APLICACIONES

7.49 De acuerdo a la ley de Maxwell-Boltzmann de la física teórica, la densidad de probabilidad de V, la velocidad de una molécula de gas, es

$$f(v) = \begin{cases} kv^2e^{-\beta v^2} & \text{para } v > 0 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

donde β depende de su masa y la temperatura absoluta y k es una constante apropiada. Demuestre que la energía cinética $E = \frac{1}{2}mV^2$ es una variable aleatoria que tiene una distribución gamma.

7.50 Con respecto al ejercicio 3.86, encuentre la densidad de probabilidad de la distancia entre el punto de impacto y el centro del blanco.

7.51 Con respecto al ejercicio 3.87, encuentre la densidad de probabilidad de la variable aleatoria $Z = \frac{X+Y}{2}$, que es el promedio de las dos proporciones de respuestas correctas que un estudiante obtendrá en las dos pruebas de aptitud.

7.52 Con respecto al ejercicio 3.88, use el teorema 7.2 para encontrar la densidad de probabilidad conjunta de las dos variables aleatorias $V = SP$ y $W = P$, y entonces encuentre la densidad marginal de V.

7.53 Use un programa de cómputo para generar 10 números "pseudo aleatorios" que tengan la distribución normal estándar.

7.54 Describa cómo quienes escribieron el software que usted utilizó para el resultado del ejercicio 7.53 pudieron haber usado la transformación de la integral de probabilidad.

7.5 TÉCNICA DE FUNCIÓN GENERATRIZ DE MOMENTOS

Las funciones generatrices de momentos pueden jugar un papel importante en la determinación de la densidad o distribución de probabilidad de una función de variables aleatorias cuando la función es una combinación lineal de n variables aleatorias *independientes*. Aquí ilustraremos esta técnica cuando una combinación lineal tal es, de hecho, la suma de n variables aleatorias independientes, le dejamos al lector que lo generalice en los ejercicios 7.59 y 7.60.

El método está basado en el teorema que la función generatriz de momentos de la suma de n variables aleatorias independientes es igual al producto de sus funciones generatrices de momentos, esto es,

TEOREMA 7.3 Si $X_1, X_2, \ldots,$ y X_n son variables aleatorias independientes y $Y = X_1 + X_2 + \cdots + X_n$, entonces

$$M_Y(t) = \prod_{i=1}^{n} M_{X_i}(t)$$

donde $M_{X_i}(t)$ es el valor de la función generatriz de momentos de X_i en t.

Demostración. Hacemos uso del hecho que las variables aleatorias son independientes y por tanto

$$f(x_1, x_2, \ldots, x_n) = f_1(x_1) \cdot f_2(x_2) \cdot \ldots \cdot f_n(x_n)$$

de acuerdo a la definición 3.14 podemos escribir

$$\begin{aligned}
M_Y(t) &= E(e^{Yt}) \\
&= E\left[e^{(X_1+X_2+\cdots+X_n)t}\right] \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{(x_1+x_2+\cdots+x_n)t} f(x_1, x_2, \ldots, x_n)\, dx_1\, dx_2 \cdots dx_n \\
&= \int_{-\infty}^{\infty} e^{x_1 t} f_1(x_1)\, dx_1 \cdot \int_{-\infty}^{\infty} e^{x_2 t} f_2(x_2)\, dx_2 \cdots \int_{-\infty}^{\infty} e^{x_n t} f_n(x_n)\, dx_n \\
&= \prod_{i=1}^{n} M_{X_i}(t)
\end{aligned}$$

lo cual demuestra el teorema para el caso continuo. Para demostrarlo para el caso discreto, sólo tenemos que reemplazar todas las integrales por sumas. ▼

Advierta que si queremos usar el teorema 7.3 para encontrar la distribución de probabilidad o la densidad de probabilidad de la variable aleatoria $Y = X_1 + X_2 + \cdots + X_n$, debemos poder identificar cualquier densidad o distribución de probabilidad que corresponda a $M_Y(t)$. Entonces, debemos basarnos en el primero de los dos teoremas que dimos en la página 224, el teorema de unicidad sobre la correspondencia entre las funciones generatrices de momentos y las densidades o distribuciones de probabilidad.

EJEMPLO 7.15

Encuentre la distribución de probabilidad de la suma de n variables aleatorias independientes $X_1, X_2, \ldots, X_n$ que tienen distribuciones de Poisson con los respectivos parámetros $\lambda_1, \lambda_2, \ldots, \lambda_n$.

Solución

Por el teorema 5.9 tenemos

$$M_{X_i}(t) = e^{\lambda_i(e^t-1)}$$

y por tanto para $Y = X_1 + X_2 + \cdots + X_n$, obtenemos

$$M_Y(t) = \prod_{i=1}^{n} e^{\lambda_i(e^t-1)} = e^{(\lambda_1+\lambda_2+\cdots+\lambda_n)(e^t-1)}$$

que se puede identificar rápidamente como la función generatriz de momentos de la distribución de Poisson con el parámetro $\lambda = \lambda_1 + \lambda_2 + \cdots + \lambda_n$. Así, la distribución de la suma de n variables aleatorias independientes que tienen distribuciones de Poisson con los parámetros λ_i es una distribución de Poisson con el parámetro $\lambda = \lambda_1 + \lambda_2 + \cdots + \lambda_n$. Observe que en el ejemplo 7.10 demostramos esto para $n = 2$. ▲

EJEMPLO 7.16

Si $X_1, X_2, \ldots, X_n$ sin variables aleatorias independientes que tienen distribuciones exponenciales con el mismo parámetro θ, encuentre la densidad de probabilidad de la variable aleatoria $Y = X_1 + X_2 + \cdots + X_n$.

Solución

Puesto que la distribución exponencial es una distribución gamma con $\alpha = 1$ y $\beta = \theta$, tenemos

$$M_{X_i}(t) = (1 - \theta t)^{-1}$$

por el teorema 6.4, y por tanto

$$M_Y(t) = \prod_{i=1}^{n} (1 - \theta t)^{-1} = (1 - \theta t)^{-n}$$

de acuerdo a la segunda de las reglas especiales para productos dadas en el apéndice A. Identificamos la función generatriz de momentos de Y como una distribución gamma con $\alpha = n$ y $\beta = \theta$, concluimos que la distribución de la suma de n variables aleatorias independientes que tienen distribuciones exponenciales con el mismo parámetro θ es una distribución gamma con los parámetros $\alpha = n$ y $\beta = \theta$. Advierta que esto concuerda con el resultado del ejemplo 7.14, donde mostramos que la suma de tres variables aleatorias independientes que tienen distribuciones exponenciales con el parámetro $\theta = 1$ tiene una distribución gamma con $\alpha = 3$ y $\beta = 1$. ▲

El teorema 7.3 también proporciona una manera fácil y elegante de derivar la función generatriz de momentos de la distribución binomial. Supongamos que X_1, $X_2, \ldots, X_n$ son variables aleatorias independientes que tienen la misma distribución Bernoulli $f(x; \theta) = \theta^x(1 - \theta)^{1-x}$ para $x = 0, 1$. Por la definición 4.6, tenemos así

$$M_{X_i}(t) = e^{0 \cdot t}(1 - \theta) + e^{1 \cdot t}\theta = 1 + \theta(e^t - 1)$$

de manera que el teorema 7.3 produce

$$M_Y(t) = \prod_{i=1}^{n} [1 + \theta(e^t - 1)] = [1 + \theta(e^t - 1)]^n$$

Esta función generatriz de momentos se identifica rápidamente como la de la distribución binomial con los parámetros n y θ. Por supuesto, $Y = X_1 + X_2 + \cdots + X_n$ es el número total de éxitos en n ensayos, puesto que X_1 es el número de éxitos en el primer intento, X_2 es el número de éxitos en el segundo intento, ..., y X_n es el número de éxitos en el $n$ésimo intento. Como veremos más adelante, esta es una manera muy fructífera de ver la distribución binomial.

EJERCICIOS

7.55 Use la técnica de la función generatriz de momentos para rehacer el ejercicio 7.35.

7.56 Encuentre la función generatriz de momentos de la distribución binomial negativa al hacer uso del hecho que si k variables aleatorias independientes tienen distribuciones geométricas con el mismo parámetro θ, su suma es una variable aleatoria que tiene la distribución binomial negativa con los parámetros θ y k. (*Sugerencia*: use el resultado del ejercicio 5.37.)

7.57 Si n variables aleatorias independientes tienen la misma distribución gamma con los parámetros α y β, encuentre la función generatriz de momentos de su suma y, si es posible, identifique su distribución.

7.58 Si n variables aleatorias independientes X_i tienen distribuciones normales con las medias μ_i y las desviaciones estándar σ_i, encuentre la función generatriz de momentos de su suma e identifique la distribución correspondiente, su media y su varianza.

7.59 Demuestre la siguiente generalización del teorema 7.3: Si $X_1, X_2, \ldots$ y X_n son variables aleatorias independientes y $Y = a_1X_1 + a_2X_2 + \cdots + a_nX_n$, entonces

$$M_Y(t) = \prod_{i=1}^{n} M_{X_i}(a_i t)$$

donde $M_{X_i}(t)$ es el valor de la función generatriz de momentos de X_i en t.

7.60 Use el resultado del ejercicio 7.59 para mostrar que, si n variables aleatorias independientes X_i tienen distribuciones normales con las medias μ_i y las desviaciones estándar σ_i, entonces $Y = a_1X_1 + a_2X_2 + \cdots + a_nX_n$ tiene una distribución normal. ¿Cuáles son la media y la desviación estándar de esta distribución?

APLICACIONES

7.61 Una abogada tiene un número privado en el cual recibe en promedio 2.1 llamadas cada media hora y un número que se encuentra en el directorio telefónico en el cual recibe en promedio 10.9 llamadas cada media hora. Si se puede suponer que los números de llamadas que recibe en estos teléfonos son variables aleatorias independientes que tienen distribuciones de Poisson, ¿cuáles son las probabilidades de que en media hora ella reciba en total

(a) 14 llamadas; (b) cuando mucho seis llamadas?

7.62 En un periódico, un vendedor de autos anuncia un Chrysler Le Baron 1993, un Ford Escort 1991, y un Buick Skylark 1994. Si el número de entrevistas que tendrá sobre estos autos se puede considerar como variables aleatorias independientes que tienen distribuciones de Poisson con los parámetros $\lambda_1 = 3.6$, $\lambda_2 = 5.8$ y $\lambda_3 = 4.6$, ¿cuáles son las probabilidades de que en total tenga

(a) menos de 10 entrevistas sobre estos autos;

(b) cualquier número entre 15 y 20 entrevistas sobre estos autos;

(c) al menos 18 entrevistas sobre estos autos?

7.63 Con respecto al ejercicio 7.62, ¿cuál es la probabilidad de que el vendedor de autos tenga seis entrevistas sobre el Ford Escort 1991 y ocho entrevistas sobre los otros dos autos?

7.64 Si el número de quejas que un negocio recibe por día es una variable aleatoria que tiene la distribución de Poisson con $\lambda = 3.3$, ¿cuáles son las probabilidades de que recibirá

(a) dos quejas en un día dado;

(b) cinco quejas en total en dos días cualquiera;

(c) al menos 12 quejas en total en tres días dados cualquiera?

7.65 El número de peces que una persona pesca por hora en Woods Canyon Lake es una variable aleatoria que tiene la distribución de Poisson con $\lambda = 1.6$. ¿Cuáles son las probabilidades de que una persona que pesca ahí atrapará

(a) cuatro peces en 2 horas;

(b) al menos dos peces en 3 horas;

(c) cuando mucho tres peces en 4 horas?

7.66 Si el número de minutos que tarda un empleado en una estación de servicio para balancear un neumático es una variable aleatoria que tiene la distribución exponencial con el parámetro $\theta = 5$, ¿cuáles son las probabilidades de que el empleado tardará

(a) menos de 8 minutos en balancear dos neumáticos;

(b) al menos 12 minutos para balancear tres neumáticos?

7.67 Si el número de minutos que un doctor dedica a un paciente es una variable aleatoria que tiene una distribución exponencial con el parámetro $\theta = 9$, ¿cuáles son las probabilidades de que el doctor tardará al menos 20 minutos para tratar

(a) un paciente; (b) dos pacientes; (c) tres pacientes?

REFERENCIAS

El uso de la transformación de la integral de probabilidad en problemas de simulación se examina en

JOHNSON, R. A., *Miller and Freund's Probability and Statistics for Engineers*, 5th ed. Upper Saddle River, N.J.: Prentice Hall, 1994.

Una generalización del teorema 7.1, el cual se aplica cuando el intervalo dentro del dominio de X para el cual $f(x) \neq 0$ se puede dividir en k subintervalos de manera que las condiciones del teorema 7.1 se aplican separadamente para cada uno de los subintervalos, se puede encontrar en

WALPOLE, R. E., and MYERS, R. H., *Probability and Statistics for Engineers and Scientists*, 4th ed. Nueva York: Macmillan Publishing Company, Inc., 1989.

En muchos textos avanzados sobre estadística matemática se dan tratamientos más detallados y más avanzados del material de este capítulo; por ejemplo, en

HOGG, R. V., and CRAIG, A. T., *Introduction to Mathematical Statistics*, 4th ed. Nueva York: Macmillan Publishing Company, Inc., 1978.

ROUSSAS, G. G., *A First Course in Mathematical Statistics*. Reading, Mass.: Addison-Wesley Publishing Company, Inc., 1973.

WILKS, S. S., *Mathematical Statistics*. Nueva York: John Wiley & Sons, Inc., 1962.

CAPÍTULO 8

Distribuciones de muestreo

8.1 INTRODUCCIÓN

La estadística se ocupa principalmente de las conclusiones y las predicciones provenientes de los resultados fortuitos que ocurren en experimentos o investigaciones cuidadosamente planeadas. En el caso finito, estos resultados fortuitos constituyen un subconjunto, o **muestra**, de las mediciones u observaciones de un conjunto mayor de valores llamado **población**. En el caso continuo generalmente son valores de variables aleatorias distribuidas en forma idéntica, a cuya distribución nos referimos como **distribución de población** o **población infinita** muestreada. La palabra "infinita" implica que, hablando en forma lógica, no hay límite al número de valores que podrían observarse.

Todos estos términos se usan aquí en forma poco convencional. Si una científica debe escoger y después pesar cinco de 40 conejillos de india como parte de un experimento, a persona sin conocimientos en la materia podría decir que los conejilos que ella selecciona constituyen una muestra. Así es como el término "muestra" se usa en el lenguaje cotidiano. En estadística, es preferible considerar los pesos de los cinco conejillos de indias como una muestra de la población, que consiste en el peso de todos los 40 conejillos de indias. En esta forma la población, así como la muestra, consiste en números. También, supongamos que, para estimar el promedio de la vida útil de cierta clase de transistor, un ingeniero selecciona 10 de estos transistores, los prueba durante cierto periodo, y registra para cada uno la hora en que se descompone. Si estas horas de descompostura son valores de variables aleatorias que tienen una distribución exponencial con parámetro θ, decimos que constituyen una muestra de esta población exponencial.

Como bien nos podemos imaginar, no todas las muestras se prestan a generalizaciones válidas sobre la población de donde vinieron. De hecho, la mayoría de los métodos de inferencia que se examinan en este libro se basan en la suposición de que estamos tratando con **muestras aleatorias.** En la práctica, a menudo tratamos con muestras aleatorias de poblaciones finitas, pero suficientemente grandes para tratarse como si fueran infinitas. Así, la mayoría de la teoría estadística y la mayoría de los métodos que examinaremos se aplican a muestras de poblaciones infinitas, y empezaremos aquí con una definición de muestras aleatorias de poblaciones infinitas. Las muestras aleatorias de poblaciones finitas se tratarán más adelante en la sección 8.3.

DEFINICIÓN 8.1 Si $X_1, X_2, \ldots, X_n$ son variables aleatorias independientes y distribuidas en forma idéntica, decimos que constituyen una **muestra aleatoria** de la población infinita dada por su distribución común.

Si $f(x_1, x_2, \ldots, x_n)$ es el valor de la distribución conjunta de un conjunto tal de variables aleatorias en $(x_1, x_2, \ldots, x_n)$, podemos escribir

$$f(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f(x_i)$$

donde $f(x_i)$ es el valor de la distribución de la población en x_i. Observe que la definición 8.1 y el análisis subsecuente se aplican sólo al muestreo con reemplazo de poblaciones finitas; el muestreo sin reemplazo de poblaciones finitas se examina en las páginas 272 y 273.

Las inferencias estadísticas suelen basarse en las **estadísticas**, esto es, en variables aleatorias que son funciones de un conjunto de variables aleatorias $X_1, X_2, \ldots, X_n$, que constituyen una muestra aleatoria. Ejemplos de lo que significa "estadísticas" son la **media de la muestra** y la **varianza de la muestra.**

DEFINICIÓN 8.2 Si $X_1, X_2, \ldots, X_n$ constituyen una muestra aleatoria entonces

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

se llama la **media de la muestra** y

$$S^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n - 1}$$

se llama la **varianza de la muestra.**†

Como se dan aquí, estas definiciones sólo se aplican a muestras aleatorias, pero la media de la muestra y la varianza de la muestra se pueden definir de la misma manera para cualquier conjunto de variables aleatorias $X_1, X_2, \ldots, X_n$.

También es común aplicar los términos "muestra aleatoria", "estadística", "media de la muestra" y "varianza de la muestra" a los valores de las variables aleatorias

† La razón de dividir entre $n - 1$ en vez de la elección aparentemente más lógica, n, se explicará en la sección 10.3.

en vez de a las variables aleatorias en sí. Intuitivamente, esto tiene más sentido y se apega al uso coloquial. Así, podríamos calcular

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \quad \text{y} \quad s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$$

para los datos observados de la muestra y referirse a estas estadísticas como la media de la muestra y la varianza de la muestra. Aquí, las x_i, $\bar{x}$ y s^2 son valores de las variables aleatorias correspondientes X_i, $\overline{X}$ y S^2. Ciertamente, las fórmulas para $\bar{x}$ y s^2 se usan aun cuando tratamos con cualquier clase de datos, no necesariamente con datos de muestreo, en cuyo caso nos referimos a $\bar{x}$ y s^2 simplemente como la media y la varianza.

Se debe entender que aquí hemos presentado $\overline{X}$ y S^2 meramente como ejemplos de estadísticas y que hay muchas otras estadísticas que se presentarán más adelante en este capítulo y en capítulos subsecuentes.

8.2 LA DISTRIBUCIÓN DE LA MEDIA

Puesto que las estadísticas son variables aleatorias, sus valores variarán de muestra a muestra, y es costumbre referirse a sus distribuciones como **distribuciones de muestreo**. La mayor parte del resto de este capítulo se dedicará a las distribuciones de muestreo de estadísticas que juegan un papel importante en las aplicaciones.

Primero estudiemos un poco de teoría sobre **la distribución de muestreo de la media**, haremos sólo algunas suposiciones muy generales sobre la naturaleza de las poblaciones muestreadas.

TEOREMA 8.1 Si $X_1, X_2, \ldots, X_n$ constituyen una muestra aleatoria de una población infinita con la media μ y la varianza σ^2, entonces

$$E(\overline{X}) = \mu \quad \text{y} \quad \text{var}(\overline{X}) = \frac{\sigma^2}{n}$$

Demostración. Hagamos $Y = \overline{X}$ en el teorema 4.14 y por tanto al hacer $a_i = \frac{1}{n}$, obtenemos

$$E(\overline{X}) = \sum_{i=1}^{n} \frac{1}{n} \cdot \mu = n\left(\frac{1}{n} \cdot \mu\right) = \mu$$

puesto que $E(X_i) = \mu$. Entonces, por el corolario del teorema 4.14, concluimos que

$$\text{var}(\overline{X}) = \sum_{i=1}^{n} \frac{1}{n^2} \cdot \sigma^2 = n\left(\frac{1}{n^2} \cdot \sigma^2\right) = \frac{\sigma^2}{n} \qquad \blacktriangledown$$

Se acostumbra escribir $E(\overline{X})$ como $\mu_{\overline{X}}$ y $\text{var}(\overline{X})$ como $\sigma^2_{\overline{X}}$ y referirse a $\sigma_{\overline{X}}$ como el **error estándar de la media**. La fórmula para el error estándar de la media,

$\sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{n}}$, muestra que la desviación estándar de la distribución de $\bar{X}$ disminuye cuando n, el **tamaño de la muestra**, se aumenta. Esto significa que cuando n se vuelve mayor y en realidad tenemos más información (los valores de más variables aleatorias), podemos esperar valores de $\bar{X}$ más cercanos a μ, la cantidad que se proponen estimar. Si nos referimos al teorema de Chebyshev como se planteó en el ejercicio 4.40, podemos expresar esto formalmente de la siguiente manera

> **TEOREMA 8.2** Para cualquier constante positiva c, la probabilidad de que $\bar{X}$ asumirá un valor entre $\mu - c$ y $\mu + c$ es cuando menos
>
> $$1 - \frac{\sigma^2}{nc^2}$$
>
> Cuando $n \to \infty$, esa probabilidad se aproxima a 1.

Este resultado, llamado **ley de números grandes**, es de interés teórico primordialmente. De mucho más valor práctico es el **teorema del límite central**, uno de los teoremas más importantes de la estadística, que tiene que ver con la distribución límite de la **media estandarizada** de n variables aleatorias cuando $n \to \infty$. Aquí sólo demostraremos este teorema para el caso donde las n variables aleatorias son una muestra aleatoria de una población cuya función generatriz de momentos existe. En los ejercicios 8.7 y 8.9 se dan condiciones más generales bajo las cuales este teorema es válido, y al final de este capítulo se hace referencia a las condiciones más generales en que este teorema es válido.

> **TEOREMA 8.3** (Teorema del límite central) Si $X_1, X_2, \ldots, X_n$ constituyen una muestra aleatoria de una población infinita con la media μ, la varianza σ^2 y la función generatriz de momentos $M_X(t)$, entonces la distribución límite de
>
> $$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$
>
> conforme $n \to \infty$ es la distribución normal estándar.

Demostración. Si primero usamos la tercera parte del teorema 4.10 y después usamos la segunda, obtenemos

$$M_Z(t) = M_{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}(t) = e^{-\sqrt{n}\,\mu t/\sigma} \cdot M_{\bar{X}}\left(\frac{\sqrt{n}\,t}{\sigma}\right)$$

$$= e^{-\sqrt{n}\,\mu t/\sigma} \cdot M_{n\bar{X}}\left(\frac{t}{\sigma\sqrt{n}}\right)$$

Puesto que $n\bar{X} = X_1 + X_2 + \cdots + X_n$, se sigue por el teorema 7.3 que

$$M_Z(t) = e^{-\sqrt{n}\,\mu t/\sigma} \cdot \left[M_X\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n$$

y por tanto que

$$\ln M_Z(t) = -\frac{\sqrt{n}\,\mu t}{\sigma} + n \cdot \ln M_X\left(\frac{t}{\sigma\sqrt{n}}\right)$$

Expandimos $M_X\left(\frac{t}{\sigma\sqrt{n}}\right)$ como una serie exponencial en t, y obtenemos

$$\ln M_Z(t) = -\frac{\sqrt{n}\,\mu t}{\sigma} + n \cdot \ln\left[1 + \mu_1' \frac{t}{\sigma\sqrt{n}} + \mu_2' \frac{t^2}{2\sigma^2 n} + \mu_3' \frac{t^3}{6\sigma^3 n\sqrt{n}} + \cdots\right]$$

donde μ_1', μ_2' y μ_3' son los momentos alrededor del origen de la distribución de población, esto es, aquellos de las variables aleatorias originales X_i.

Si n es lo suficientemente grande, podemos usar la expansión de $\ln(1 + x)$ como una serie exponencial en x (como en la página 223), y obtenemos

$$\begin{aligned}\ln M_Z(t) = -\frac{\sqrt{n}\,\mu t}{\sigma} + n\Bigg\{&\left[\mu_1' \frac{t}{\sigma\sqrt{n}} + \mu_2' \frac{t^2}{2\sigma^2 n} + \mu_3' \frac{t^3}{6\sigma^3 n\sqrt{n}} + \cdots\right] \\ &- \frac{1}{2}\left[\mu_1' \frac{t}{\sigma\sqrt{n}} + \mu_2' \frac{t^2}{2\sigma^2 n} + \mu_3' \frac{t^3}{6\sigma^3 n\sqrt{n}} + \cdots\right]^2 \\ &+ \frac{1}{3}\left[\mu_1' \frac{t}{\sigma\sqrt{n}} + \mu_2' \frac{t^2}{2\sigma^2 n} + \mu_3' \frac{t^3}{6\sigma^3 n\sqrt{n}} + \cdots\right]^3 - \cdots\Bigg\}\end{aligned}$$

Después, reunimos las potencias de t, y obtenemos

$$\begin{aligned}\ln M_Z(t) = &\left(-\frac{\sqrt{n}\,\mu}{\sigma} + \frac{\sqrt{n}\,\mu_1'}{\sigma}\right)t + \left(\frac{\mu_2'}{2\sigma^2} - \frac{\mu_1'^2}{2\sigma^2}\right)t^2 \\ &+ \left(\frac{\mu_3'}{6\sigma^3\sqrt{n}} - \frac{\mu_1' \cdot \mu_2'}{2\sigma^3\sqrt{n}} + \frac{\mu_1'^3}{3\sigma^3\sqrt{n}}\right)t^3 + \cdots\end{aligned}$$

y puesto que $\mu_1' = \mu$ y $\mu_2' - \mu_1' = \sigma^2$, esto se simplifica a

$$\ln M_Z(t) = \frac{1}{2}t^2 + \left(\frac{\mu_3'}{6} - \frac{\mu_1'\mu_2'}{2} + \frac{\mu_1'^3}{6}\right)\frac{t^3}{\sigma^3\sqrt{n}} + \cdots$$

Finalmente, al observar que el coeficiente t^3 es una constante multiplicada por $\frac{1}{\sqrt{n}}$ y en general, para $r \geqq 2$, el coeficiente de t^r es una constante multiplicada por $\frac{1}{\sqrt{n^{r-2}}}$, obtenemos

$$\lim_{n\to\infty} \ln M_Z(t) = \frac{1}{2}t^2$$

y por tanto

$$\lim_{n\to\infty} M_Z(t) = e^{\frac{1}{2}t^2}$$

puesto que el límite de un logaritmo es igual al logaritmo del límite (a condición de que estos límites existan). Identificamos la función generatriz de momentos a la que hemos llegado como la de la distribución normal estándar, sólo necesitamos los dos teoremas enunciados en la página 224 para completar la demostración del teorema 8.3. ▼

Algunas veces, el teorema del límite central se interpreta incorrectamente como si implicara que la distribución de $\overline{X}$ se aproxima a una distribución normal cuando $n \to \infty$. Esto es incorrecto porque $\text{var}(\overline{X}) \to 0$ cuando $n \to \infty$; por otra parte, el teorema del límite central justifica la aproximación de la distribución de $\overline{X}$ con una distribución normal que tenga una media μ y la varianza $\frac{\sigma^2}{n}$ cuando n es grande. En la práctica, esta aproximación se usa cuando $n \geqq 30$ sin importar la forma real de la población muestreada. Para valores menores de n la aproximación es dudosa, pero vea el teorema 8.4.

EJEMPLO 8.1

Una máquina de refrescos está arreglada para que la cantidad de bebida que sirve sea una variable aleatoria con una media de 200 mililitros y una desviación estándar de 15 mililitros. ¿Cuál es la probabilidad de que la cantidad promedio (media) servida en una muestra aleatoria de tamaño 36 sea al menos de 204 mililitros?

Solución

De acuerdo al teorema 8.1, la distribución de $\overline{X}$ tiene la media $\mu_{\overline{X}} = 200$ y la desviación estándar $\sigma_{\overline{X}} = \frac{15}{\sqrt{36}} = 2.5$, y de acuerdo al teorema del límite central, esta distribución es aproximadamente normal. Puesto que $z = \frac{204 - 200}{2.5} = 1.6$, se sigue de la tabla III que $P(\overline{X} \geqq 204) = P(Z \geqq 1.6) = 0.5000 - 0.4452 = 0.0548$. ▲

Es interesante advertir que cuando la población que estamos muestreando es normal, la distribución de $\overline{X}$ es una distribución normal sin importar el tamaño de n.

TEOREMA 8.4 Si $\overline{X}$ es la media de una muestra aleatoria de tamaño n de una población normal con la media μ y la varianza σ^2, su distribución de muestreo es una distribución normal con la media μ y la varianza σ^2/n.

Demostración. De acuerdo con los teoremas 4.10 y 7.3, podemos escribir

$$M_{\overline{X}}(t) = \left[M_X\left(\frac{t}{n}\right)\right]^n$$

y puesto que la función generatriz de momentos de una distribución normal con la media μ y la varianza σ^2 está dada por

$$M_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$$

de acuerdo al teorema 6.6, obtenemos

$$\begin{aligned} M_{\bar{X}}(t) &= \left[e^{\mu \cdot \frac{t}{n} + \frac{1}{2}\left(\frac{t}{n}\right)^2 \sigma^2} \right]^n \\ &= e^{\mu t + \frac{1}{2} t^2 \left(\frac{\sigma^2}{n}\right)} \end{aligned}$$

Esta función generatriz de momentos se ve en seguida que es la de una distribución normal con la media μ y la varianza σ^2/n, y para completar la demostración del teorema 8.4 sólo tenemos que referirnos a los dos teoremas de la página 224. ▼

8.3 LA DISTRIBUCIÓN DE LA MEDIA: POBLACIONES FINITAS

Si un experimento consiste en seleccionar uno o más valores de un conjunto finito de números $\{c_1, c_2, \dots, c_N\}$, este conjunto se conoce como **población finita de tamaño** N. En la definición que sigue se supondrá que estamos muestreando sin reemplazo de una población finita de tamaño N.

> **DEFINICIÓN 8.3** Si X_1 es el primer valor sacado de una población finita de tamaño N, X_2 es el segundo valor sacado, ..., X_n es el $n$ésimo valor sacado, y la distribución de probabilidad conjunta de estas n variables aleatorias está dada por
>
> $$f(x_1, x_2, \dots, x_n) = \frac{1}{N(N-1)\cdot \ldots \cdot(N-n+1)}$$
>
> para cada $n$tuplo ordenado de valores de estas variables aleatorias, entonces se dice que X_1, X_2 y X_n constituyen una **muestra aleatoria** de la población finita dada.

Como en la definición 8.1, la muestra aleatoria es un conjunto de variables aleatorias, pero, aquí una vez más, también es común aplicar el término "muestra aleatoria" a los valores de las variables aleatorias, esto es, a los números reales extraídos.

De la distribución de probabilidad conjunta de la definición 8.3, se sigue que la probabilidad de cada subconjunto de n de estos N elementos de la población finita (sin importar el orden en el cual se saquen los valores) es

$$\frac{n!}{N(N-1)\cdot \ldots \cdot(N-n+1)} = \frac{1}{\binom{N}{n}}$$

A menudo esto se da como una definición alternativa o como un criterio para la selección de una muestra aleatoria de tamaño n de una población finita de tamaño N: *cada una de las* $\binom{N}{n}$ *muestras posibles debe tener la misma probabilidad.*

También se sigue de la definición 8.3 de la distribución de probabilidad conjunta que la distribución marginal de X_r está dada por

$$f(x_r) = \frac{1}{N} \qquad \text{para } x_r = c_1, c_2, \ldots, c_N$$

para $r = 1, 2, \ldots, n$, y nos referimos a la media y la varianza de esta distribución uniforme discreta como la media y la varianza de la población finita. Por consiguiente,

DEFINICIÓN 8.4 La media y la **varianza** de la población finita $\{c_1, c_2, \ldots, c_N\}$ son

$$\mu = \sum_{i=1}^{N} c_i \cdot \frac{1}{N} \quad \text{y} \quad \sigma^2 = \sum_{i=1}^{N} (c_i - \mu)^2 \cdot \frac{1}{N}$$

Finalmente, de la definición 8.3 de la distribución de probabilidad conjunta se sigue que la distribución marginal conjunta de dos variables aleatorias cualesquiera $X_1, X_2, \ldots, X_n$ está dada por

$$g(x_r, x_s) = \frac{1}{N(N-1)}$$

para cada par ordenado de elementos de la población finita. Así, podemos demostrar que

TEOREMA 8.5 Si X_r y X_s son las $r$ésima y el $s$ésima variables aleatorias de una muestra aleatoria de tamaño n sacada de una población finita $\{c_1, c_2, \ldots, c_N\}$, entonces

$$\text{cov}(X_r, X_s) = -\frac{\sigma^2}{N-1}$$

Demostración. De acuerdo con la definición 4.9

$$\begin{aligned}\text{cov}(X_r, X_s) &= \sum_{i=1}^{N} \sum_{\substack{j=1 \\ i \neq j}}^{N} \frac{1}{N(N-1)} (c_i - \mu)(c_j - \mu) \\ &= \frac{1}{N(N-1)} \cdot \sum_{i=1}^{N} (c_i - \mu) \left[\sum_{\substack{j=1 \\ j \neq i}}^{N} (c_j - \mu) \right]\end{aligned}$$

y puesto que $\sum_{\substack{j=1 \\ j \neq i}}^{N} (c_j - \mu) = \sum_{j=1}^{N} (c_j - \mu) - (c_i - \mu) = -(c_i - \mu)$, obtenemos

$$\operatorname{cov}(X_r, X_s) = -\frac{1}{N(N-1)} \cdot \sum_{i=1}^{N} (c_i - \mu)^2$$
$$= -\frac{1}{N-1} \cdot \sigma^2 \qquad \blacktriangledown$$

Al hacer uso de todos estos resultados, demostremos ahora el siguiente teorema, el cual corresponde al teorema 8.1 para muestras aleatorias de poblaciones finitas.

TEOREMA 8.6 Si $\overline{X}$ es la media de una muestra aleatoria de tamaño n de una población finita de tamaño N con la media μ y la varianza σ^2, entonces

$$E(\overline{X}) = \mu \quad \text{y} \quad \operatorname{var}(\overline{X}) = \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1}$$

Demostración. Sustituimos $a_i = \frac{1}{N}$, $\operatorname{var}(X_i) = \sigma^2$, y $\operatorname{cov}(X_i, X_j) = -\frac{\sigma^2}{N-1}$ en la fórmula del teorema 4.14, y obtenemos

$$E(\overline{X}) = \sum_{i=1}^{n} \frac{1}{n} \cdot \mu = \mu$$

y

$$\operatorname{var}(\overline{X}) = \sum_{i=1}^{n} \frac{1}{n^2} \cdot \sigma^2 + 2 \cdot \sum\sum_{i<j} \frac{1}{n^2}\left(-\frac{\sigma^2}{N-1}\right)$$
$$= \frac{\sigma^2}{n} + 2 \cdot \frac{n(n-1)}{2} \cdot \frac{1}{n^2}\left(-\frac{\sigma^2}{N-1}\right)$$
$$= \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1} \qquad \blacktriangledown$$

Es interesante observar que las fórmulas que obtuvimos para $\operatorname{var}(\overline{X})$ en los teoremas 8.1 y 8.6 sólo difieren por el **factor de corrección por población finita** $\frac{N-n}{N-1}$.† Ciertamente cuando N es grande comparado con n, la diferencia entre las dos fórmulas para $\operatorname{var}(\overline{X})$ suele ser despreciable y la fórmula $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$ se usa a menudo como

† Puesto que hay muchos problemas que nos interesan en la desviación estándar más que en la varianza, el término "factor de corrección por población finita" a menudo se refiere a $\sqrt{\frac{N-n}{N-1}}$ en vez de $\frac{N-n}{N-1}$. Por supuesto, esto no importa en tanto el uso se entienda claramente.

una aproximación cuando estamos muestreando a partir de una población finita grande. Una regla empírica general es usar esta aproximación cuando la muestra no constituye más del 5% de la población.

EJERCICIOS

8.1 Use el corolario del teorema 4.15 para mostrar que si $X_1, X_2, \ldots, X_n$ constituyen una muestra aleatoria de una población infinita, entonces

$$\operatorname{cov}(X_r - \bar{X}, \bar{X}) = 0$$

para $r = 1, 2, \ldots, n$.

8.2 Use el teorema 4.14 y su corolario para mostrar que si $X_{11}, X_{12}, \ldots, X_{1n_1}, X_{21}, X_{22}, \ldots, X_{2n_2}$ son variables aleatorias independientes, donde la primeras n_1 constituyen una muestra aleatoria de una población infinita con la media μ_1 y la varianza σ_1^2 y las otra n_2 constituyen una muestra aleatoria de una población infinita con la media μ_2 y la varianza σ_2^2, entonces

(a) $E(\bar{X}_1 - \bar{X}_2) = \mu_1 - \mu_2$;

(b) $\operatorname{var}(\bar{X}_1 - \bar{X}_2) = \dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}$.

8.3 Con respecto al ejercicio 8.2, demuestre que si las dos muestras vienen de poblaciones normales, entonces $\bar{X}_1 - \bar{X}_2$ es una variable aleatoria que tiene la distribución normal con la media $\mu_1 - \mu_2$ y la varianza $\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}$. (*Sugerencia*: proceda como en la demostración del teorema 8.4.)

8.4 Si $X_1, X_2, \ldots, X_n$ son variables aleatorias independientes que tienen distribuciones de Bernoulli idénticas con el parámetro θ, entonces $\bar{X}$ es la proporción de éxitos en n intentos, lo que denotamos con $\hat{\Theta}$. Verifique que

(a) $E(\hat{\Theta}) = \theta$;

(b) $\operatorname{var}(\hat{\Theta}) = \dfrac{\theta(1-\theta)}{n}$.

8.5 Si las primeras n_1 variables aleatorias del ejercicio 8.2 tienen distribuciones de Bernoulli con el parámetro θ_1 y las otras n_2 variables aleatorias tienen distribuciones de Bernoulli con el parámetro θ_2, demuestre que, en la notación del ejercicio 8.4,

(a) $E(\hat{\Theta}_1 - \hat{\Theta}_2) = \theta_1 - \theta_2$;

(b) $\operatorname{var}(\hat{\Theta}_1 - \hat{\Theta}_2) = \dfrac{\theta_1(1-\theta_1)}{n_1} + \dfrac{\theta_2(1-\theta_2)}{n_2}$.

8.6 Demuestre el teorema 6.8, considere las variables aleatorias binomiales como en la página 263, esto es, como sumas de variables aleatorias de Bernoulli independientes distribuidas en forma idéntica y utilice el teorema del límite central.

8.7 Lo siguiente es una condición suficiente para el teorema del límite central: si las variables aleatorias $X_1, X_2, \ldots, X_n$ son independientes y uniformemente limitadas (esto es, existe una constante positiva k tal que la probabilidad es cero de que una de las variables aleatorias cualesquiera X_i asumirá un valor mayor que k o menor que $-k$), entonces si la varianza de

$$Y_n = X_1 + X_2 + \cdots + X_n$$

se vuelve infinita cuando $n \to \infty$, la distribución de la media estandarizada de X_i se aproxima a la distribución normal estándar. Muestre que esta condición suficiente es válida para una serie de variables aleatorias independientes X_i que tienen las respectivas distribuciones de probabilidad

$$f(x_i) = \begin{cases} \frac{1}{2} & \text{para } x_i = 1 - \left(\frac{1}{2}\right)^i \\ \frac{1}{2} & \text{para } x_i = \left(\frac{1}{2}\right)^i - 1 \end{cases}$$

8.8 Considere la serie de variables aleatorias independientes $X_1, X_2, X_3, \ldots$ que tienen las densidades uniformes

$$f(x_i) = \begin{cases} \dfrac{1}{2 - \frac{1}{i}} & \text{para } 0 < x_i < 2 - \frac{1}{i} \\ 0 & \text{en cualquier otra parte} \end{cases}$$

Use la condición suficiente del ejercicio 8.7 para mostrar que el teorema del límite central es válido.

8.9 La siguiente es una condición suficiente, la *condición Laplace-Liapounoff*, para el teorema del límite central: si $X_1, X_2, X_3, \ldots$ es una serie de variables aleatorias independientes, cada una de las cuales tiene un tercer momento absoluto

$$c_i = E(|X_i - \mu_i|^3)$$

y si

$$\lim_{n \to \infty} [\operatorname{var}(Y_n)]^{-\frac{3}{2}} \cdot \sum_{i=1}^{n} c_i = 0$$

donde $Y_n = X_1 + X_2 + \cdots + X_n$, entonces la distribución de la media estandarizada de las X_i se aproxima a la distribución normal estándar cuando $n \to \infty$. Use esta condición para demostrar que el teorema del límite central es válido para la serie de variables aleatorias del ejercicio 8.7.

8.10 Use la condición del ejercicio 8.9 para mostrar que el teorema del límite central es válido para la serie de variables aleatorias del ejercicio 8.8.

8.11 Explique por qué, cuando muestreamos con reemplazo de una población finita, se aplican los resultados del teorema 8.1 en vez de los del teorema 8.6.

8.12 Explique los resultados del ejercicio 5.45 a la luz del teorema 8.6.

8.13 Si una muestra aleatoria de tamaño n se selecciona de la población finita que consiste de los enteros $1, 2, \ldots, N$, demuestre que:

(a) la media de $\overline{X}$ es $\dfrac{N+1}{2}$;

(b) la varianza de $\overline{X}$ es $\dfrac{(N+1)(N-n)}{12n}$;

(c) la media y la varianza de $Y = n \cdot \overline{X}$ son

$$E(Y) = \frac{n(N+1)}{2} \quad \text{y} \quad \text{var}(Y) = \frac{n(N+1)(N-n)}{12}$$

(*Sugerencia*: consulte el apéndice A o los resultados del ejercicio 5.1.)

8.14 Encuentre la media y la varianza de la población finita que consiste de los 10 números 15, 13, 18, 10, 6, 21, 7, 11, 20 y 9.

8.15 Demuestre que la varianza de la población finita $\{c_1, c_2, \ldots, c_N\}$ se puede escribir como

$$\sigma^2 = \frac{\sum_{i=1}^{N} c_i^2}{N} - \mu^2$$

También, use esta fórmula para recalcular la varianza de la población finita del ejercicio 8.14.

8.16 Muestre que, análoga a la fórmula del ejercicio 8.15, la fórmula para la varianza de la muestra se puede escribir como

$$S^2 = \frac{\sum_{i=1}^{n} X_i^2}{n-1} - \frac{n\overline{X}^2}{n-1}$$

También, use esta fórmula para calcular la varianza de los siguientes datos de muestreo sobre el número de llamadas de servicio que recibe el operador de un camión grúa en ocho días de trabajo consecutivos: 13, 14, 13, 11, 15, 14, 17 y 11.

8.17 Demuestre que la fórmula para la varianza de la muestra se puede escribir como

$$S^2 = \frac{n\left(\sum_{i=1}^{n} X_i^2\right) - \left(\sum_{i=1}^{n} X_i\right)^2}{n(n-1)}$$

También, use esta fórmula para recalcular la varianza de los datos de la muestra del ejercicio 8.16.

APLICACIONES

8.18 ¿Cuántas muestras diferentes de tamaño $n = 3$ se pueden sacar de una población finita de tamaño

(a) $N = 12$; (b) $N = 20$; (c) $N = 50$?

8.19 ¿Cuál es la probabilidad de cada muestra posible si

(a) se va a sacar una muestra aleatoria de tamaño $n = 4$ de una población finita de tamaño $N = 12$;

(b) se va a sacar una muestra aleatoria de tamaño $n = 5$ de una población finita de tamaño $N = 22$?

8.20 Si se saca una muestra aleatoria de tamaño $n = 3$ de una población finita de tamaño $N = 50$, ¿cuál es la probabilidad de que un elemento en particular de la población se incluirá en la muestra?

8.21 Para muestras aleatorias de una población infinita, ¿qué pasa con el error estándar de la media si el tamaño de la muestra se

(a) aumenta de 30 a 120;

(b) aumenta de 80 a 180;

(c) disminuye de 450 a 50;

(d) disminuye de 250 a 40?

8.22 Encuentre el valor del factor de corrección por población finita $\frac{N-n}{N-1}$ para

(a) $n = 5$ y $N = 200$;

(b) $n = 50$ y $N = 300$;

(c) $n = 200$ y $N = 800$.

8.23 Se toma una muestra aleatoria de tamaño $n = 100$ de una población infinita con la media $\mu = 75$ y la varianza $\sigma^2 = 256$.

(a) Con base en el teorema de Chebyshev, ¿con qué probabilidad podemos afirmar que el valor que obtenemos para $\overline{X}$ caerá entre 67 y 83?

(b) Con base en el teorema del límite central, ¿con qué probabilidad podemos afirmar que el valor que obtenemos para $\overline{X}$ caerá entre 67 y 83?

8.24 Una muestra aleatoria de tamaño $n = 81$ se toma de una población infinita con la media $\mu = 128$ y la desviación estándar $\sigma = 6.3$. ¿Con qué probabilidad podemos afirmar que el valor que obtenemos para $\overline{X}$ no caerá entre 126.6 y 129.4 si usamos

(a) el teorema de Chebyshev;

(b) el teorema del límite central?

8.25 Vuelva a resolver el inciso (b) del ejercicio 8.24, suponiendo que la población no es infinita sino finita y de tamaño $N = 400$.

8.26 Se tomará una muestra aleatoria de tamaño $n = 225$ de una población exponencial con $\theta = 4$. Con base en el teorema del límite central, ¿cuál es la probabilidad de que la media de la muestra excederá 4.5?

8.27 Se tomará una muestra aleatoria de tamaño $n = 200$ de una población uniforme con $\alpha = 24$ y $\beta = 48$. Basado en el teorema del límite central, ¿cuál es la probabilidad de que la media de la muestra será menor que 35?

8.28 Se toma una muestra aleatoria de tamaño 64 de una población normal con $\mu = 51.4$ y $\sigma = 6.8$. ¿Cuál es la probabilidad de que la media de la muestra

(a) excederá 52.9;

(b) caerá entre 50.5 y 52.3;

(c) será menor que 50.6?

8.29 Se toma una muestra aleatoria de tamaño 100 de una población normal con $\sigma = 25$. ¿Cuál es la probabilidad de que la media de la muestra diferirá de la media de la población por 3 o más en cualquier dirección?

8.30 Se toman muestras aleatorias independientes de tamaño 400 de cada una de dos poblaciones que tienen medias iguales y desviaciones estándar $\sigma_1 = 20$ y $\sigma_2 = 30$. Al usar el teorema de Chebyshev y el resultado del ejercicio 8.2, ¿qué podemos afirmar con una probabilidad de al menos 0.99 sobre el valor que obtendremos para $\bar{X}_1 - \bar{X}_2$? (Por "independiente" queremos decir que las muestras satisfacen las condiciones del ejercicio 8.2.)

8.31 Suponga que las dos poblaciones del ejercicio 8.30 son normales y use el resultado del ejercicio 8.3 para encontrar k tal que

$$P(-k < \bar{X}_1 - \bar{X}_2 < k) = 0.99$$

8.32 Se toman muestras aleatorias independientes tamaño $n_1 = 30$ y $n_2 = 50$ de dos poblaciones normales que tienen las medias $\mu_1 = 78$ y $\mu_2 = 75$ y las varianzas $\sigma_1^2 = 150$ y $\sigma_2^2 = 200$. Use los resultados del ejercicio 8.3 para encontrar la probabilidad de que la media de la primera muestra excederá la de la segunda muestra por al menos 4.8.

8.33 La proporción real de familias que en cierta ciudad son dueñas de sus casas, en vez de rentarlas, es 0.70. Si se entrevistan aleatoriamente 84 familias en esta ciudad y sus respuestas a la pregunta de si son dueñas de sus casas se consideran como valores de variables aleatorias independientes que tienen distribuciones de Bernoulli idénticas con el parámetro $\theta = 0.70$, ¿con qué probabilidad podemos afirmar que el valor que obtenemos para la muestra de la proporción $\hat{\Theta}$ caerá entre 0.64 y 0.76, use el resultado del ejercicio 8.4 y

(a) el teorema de Chebyshev;

(b) el teorema del límite central?

8.34 La proporción real de hombres que favorecen cierta propuesta de impuestos es 0.40 y la proporción correspondiente de mujeres es 0.25; aleatoriamente se entrevistan $n_1 = 500$ hombres y $n_2 = 400$ mujeres y sus respuestas individuales se consideran como los valores de variables aleatorias independientes que tienen distribuciones de Bernoulli con los respectivos parámetros $\theta_1 = 0.40$ y $\theta_2 = 0.25$. ¿Qué podemos afirmar, de acuerdo al teorema de Chebyshev, con una probabilidad de al menos 0.9375 sobre el valor que obtendremos para $\hat{\Theta}_1 - \hat{\Theta}_2$, la diferencia entre las dos muestras de proporciones de respuestas favorables? Use el resultado del ejercicio 8.5.

8.4 LA DISTRIBUCIÓN JI CUADRADA

En el ejemplo 7.9 demostramos que si X tiene la distribución normal estándar, entonces X^2 tiene la distribución gamma especial, a la cual nos referimos como **distribución ji cuadrada**, y esto explica el papel importante que la distribución ji cuadrada tiene en problemas de muestreo de poblaciones normales. La distribución ji cuadrada a menudo se denota como "distribución χ^2" donde χ es la letra minúscula griega *ji*. También usamos χ^2 para los valores de las variables aleatorias que tienen distribuciones ji cuadrada, pero nos abstendremos de denotar las variables aleatorias correspondientes con X^2, donde X es la letra mayúscula griega *ji*. Esto evita tener que reiterar en cada caso si X es una variable aleatoria con valores x o una variable aleatoria con valores χ.

Revisemos algunos de los resultados de la sección 6.3, una variable aleatoria X tiene la distribución ji cuadrada con ν grados de libertad si su densidad de probabilidad está dada por

$$f(x) = \begin{cases} \dfrac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{\frac{\nu-2}{2}} e^{-x/2} & \text{para } x > 0 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

La media y la varianza de la distribución ji cuadrada con ν grados de libertad son ν y 2ν, y su función generatriz de momentos está dada por

$$M_X(t) = (1 - 2t)^{-\nu/2}$$

La distribución ji cuadrada tiene varias propiedades matemáticas importantes, que se dan en los teoremas 8.7 a 8.10. Primero, enunciemos formalmente el resultado del ejemplo 7.9, al cual nos referimos anteriormente

TEOREMA 8.7 Si X tiene la distribución normal estándar, entonces X^2 tiene la distribución ji cuadrada con $\nu = 1$ grados de libertad.

Más generalmente, demostremos que

TEOREMA 8.8 Si $X_1, X_2, \dots, X_n$ son variables aleatorias independientes que tienen distribuciones normal estándar, entonces

$$Y = \sum_{i=1}^{n} X_i^2$$

tiene la distribución ji cuadrada con $\nu = n$ grados de libertad.

Demostración. Al usar la función generatriz de momentos dada arriba con $\nu = 1$ y el teorema 8.7, encontramos que

$$M_{X_i^2}(t) = (1 - 2t)^{-\frac{1}{2}}$$

y se sigue, por el teorema 7.3, que

$$M_Y(t) = \prod_{i=1}^{n} (1 - 2t)^{-\frac{1}{2}} = (1 - 2t)^{-\frac{n}{2}}$$

Esta función generatriz de momentos se identifica en seguida como la de la distribución ji cuadrada con $\nu = n$ grados de libertad. ▼

Dos propiedades adicionales de la distribución ji cuadrada se dan en los dos teoremas que siguen; se pedirá al lector que los demuestre en los ejercicios 8.35 y 8.36.

TEOREMA 8.9 Si $X_1, X_2, \ldots, X_n$ son variables aleatorias independientes que tienen distribuciones ji cuadrada con $\nu_1, \nu_2, \ldots, \nu_n$ grados de libertad, entonces

$$Y = \sum_{i=1}^{n} X_i$$

tiene la distribución ji cuadrada con $\nu_1 + \nu_2 + \cdots + \nu_n$ grados de libertad.

TEOREMA 8.10 Si X_1 y X_2 son variables aleatorias independientes, X_1 tiene la distribución ji cuadrada con ν_1 grados de libertad, y $X_1 + X_2$ tiene la distribución ji cuadrada con $\nu > \nu_1$ grados de libertad, entonces X_2 tiene la distribución ji cuadrada con $\nu - \nu_1$ grados de libertad.

La distribución ji cuadrada tiene muchas aplicaciones importantes, varias de las cuales se examinan en los capítulos 10 al 13. Las más destacadas, son aquellas basadas directa o indirectamente en el siguiente teorema.

TEOREMA 8.11 Si $\overline{X}$ y S^2 son la media y la varianza de una muestra aleatoria de tamaño n de una población normal con la media μ y la desviación estándar σ, entonces

1. $\overline{X}$ y S^2 son independientes;
2. La variable aleatoria $\dfrac{(n-1)S^2}{\sigma^2}$ tiene la distribución ji cuadrada con $n - 1$ grados de libertad.

Demostración. Puesto que una demostración detallada de la parte 1 rebasaría el alcance de este libro, supondremos la independencia de $\overline{X}$ y S^2 en nuestra demostración de la parte 2. Además de las referencias a las demostraciones de la parte 1 al final de este capítulo, el ejercicio 8.46 bosqueja los pasos principales de una demostración un poco más simple basada en la idea de una función generatriz de momentos condicional, y el ejercicio 8.45 pedirá al lector que demuestre la independencia de $\overline{X}$ y S^2 para el caso especial donde $n = 2$.

Para demostrar la parte 2, empezamos con la identidad

$$\sum_{i=1}^{n} (X_i - \mu)^2 = \sum_{i=1}^{n} (X_i - \overline{X})^2 + n(\overline{X} - \mu)^2$$

que se pedirá al lector verifique en el ejercicio 8.37. Ahora, si dividimos cada término entre σ^2 y sustituimos $(n - 1)S^2$ en vez de $\sum_{i=1}^{n} (X_i - \overline{X})^2$, se sigue que

$$\sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2 = \frac{(n-1)S^2}{\sigma^2} + \left(\frac{\overline{X} - \mu}{\sigma/\sqrt{n}}\right)^2$$

En cuanto a los tres términos de esta identidad, sabemos por el teorema 8.8 que la que está en el lado izquierdo de la ecuación es una variable aleatoria que tiene la distribución ji cuadrada con n grados de libertad. También, según los teoremas 8.4 y 8.7, el segundo término en el lado derecho de la ecuación es una variable aleatoria que tiene la distribución ji cuadrada con 1 grado de libertad. Ahora, puesto que se supone que $\overline{X}$ y S^2 son independientes, se sigue que los dos términos en el lado derecho de la ecuación son independientes, y concluimos por el teorema 8.10 que $\frac{(n-1)S^2}{\sigma^2}$ es una variable aleatoria independiente que tiene una distribución ji cuadrada con $n-1$ grados de libertad. ▼

Puesto que la distribución ji cuadrada se presenta en muchas aplicaciones importantes, se han tabulado extensivamente las integrales de su densidad. La tabla V contiene los valores de $\chi^2_{\alpha,\nu}$ para $\alpha = 0.995, 0.99, 0.975, 0.95, 0.05, 0.025, 0.01$ y 0.005, y $\nu = 1, 2, \ldots, 30$, donde $\chi^2_{\alpha,\nu}$ es tal que el área a su derecha bajo la curva ji cuadrada con ν grados de libertad (véase la figura 8.1) es igual a α. Esto es, $\chi^2_{\alpha,\nu}$ es tal que si X es una variable aleatoria que tiene la distribución ji cuadrada con ν grados de libertad, entonces

$$P(X \geqq \chi^2_{\alpha,\nu}) = \alpha$$

Cuando ν es mayor que 30, no se puede usar la tabla V y las probabilidades referentes a las distribuciones ji cuadrada usualmente se aproximan con las distribuciones normales, como en el ejercicio 8.40 u 8.43.

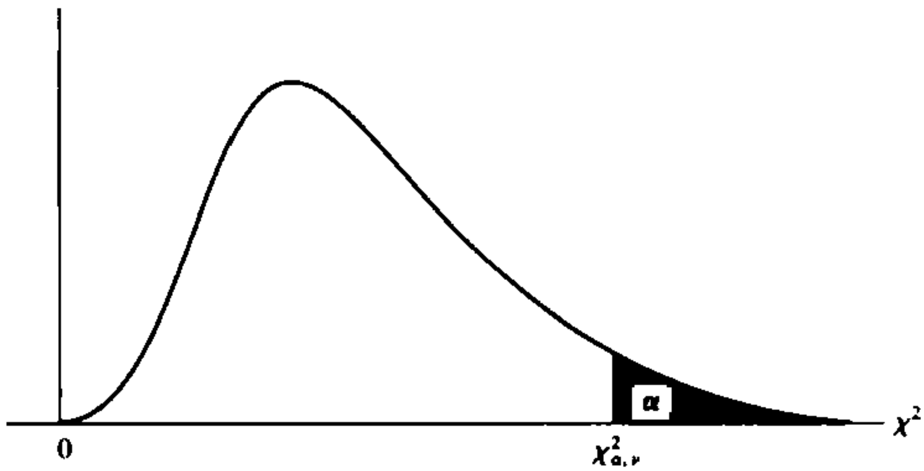

Figura 8.1 Distribución ji cuadrada.

EJEMPLO 8.2

Supongamos que el espesor de una parte usada en un semiconductor es su dimensión crítica y que el proceso de fabricar estas partes se considera que está bajo control si la variación real entre el espesor de las partes está dada por una desviación estándar no mayor que $\sigma = 0.60$ milésimas de pulgada. Para mantener un control sobre el proceso, periódicamente se toman muestras aleatorias de tamaño $n = 20$ y se considera que está "fuera de control" si la probabilidad de que S^2 asumirá un valor mayor que, o igual, al valor observado de la muestra es 0.01 o menor (aun cuando $\sigma = 0.60$). ¿Qué puede uno concluir sobre el proceso si la desviación estándar de una muestra aleatoria periódica tal es $s = 0.84$ milésimas de pulgada?

Solución

El proceso se declarará "fuera de control" si $\frac{(n-1)s^2}{\sigma^2}$ con $n = 20$ y $\sigma = 0.60$ excede a $\chi^2_{0.01,\,19} = 36.191$. Puesto que

$$\frac{(n-1)s^2}{\sigma^2} = \frac{19(0.84)^2}{(0.60)^2} = 37.24$$

excede 36.191, se declara el proceso fuera de control. Por supuesto, se supone que la muestra se puede considerar una muestra aleatoria de una población normal. ▲

8.5 LA DISTRIBUCIÓN t

En el teorema 8.4 demostramos que para muestras aleatorias de una población normal con la media μ y la varianza σ^2, la variable aleatoria $\overline{X}$ tiene una distribución normal con la media μ y la varianza $\frac{\sigma^2}{n}$; en otras palabras,

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

tiene la distribución normal estándar. Éste es un resultado importante, pero la mayor dificultad para aplicarlo es que en las aplicaciones más reales se desconoce la desviación estándar de la población, σ. Esto hace necesario reemplazar σ con un estimado, usualmente con el valor de la desviación estándar de la muestra S. Así, la teoría que sigue nos lleva a la distribución exacta de $\frac{\overline{X} - \mu}{S/\sqrt{n}}$ para muestras aleatorias de poblaciones normales.

Para derivar esta distribución de muestreo, estudiemos primero la situación más general tratada en el teorema siguiente.

TEOREMA 8.12 Si Y y Z son variables aleatorias independientes, Y tiene la distribución ji cuadrada con ν grados de libertad, y Z tiene la distribución normal estándar, entonces la distribución de

$$T = \frac{Z}{\sqrt{Y/\nu}}$$

está dada por

$$f(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\,\Gamma\left(\frac{\nu}{2}\right)} \cdot \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} \qquad \text{para } -\infty < t < \infty$$

y se llama la **distribución t** con ν grados de libertad.

Demostración. Puesto que Y y Z son independientes, su densidad de probabilidad conjunta está dada por

$$f(y, z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \cdot \frac{1}{\Gamma\left(\frac{\nu}{2}\right)2^{\frac{\nu}{2}}} y^{\frac{\nu}{2}-1} e^{-\frac{y}{2}}$$

para $y > 0$ y $-\infty < z < \infty$, y $f(y, z) = 0$ en cualquier otra parte. Entonces, para usar la técnica de cambio de variable de la sección 7.3, resolvemos $t = \frac{z}{\sqrt{y/\nu}}$ para z, y obtenemos $z = t\sqrt{y/\nu}$ y por tanto $\frac{\partial z}{\partial t} = \sqrt{y/\nu}$. Así, por el teorema 7.1, la densidad conjunta de Y y T está dada por

$$g(y, t) = \begin{cases} \dfrac{1}{\sqrt{2\pi\nu}\,\Gamma\left(\frac{\nu}{2}\right)2^{\frac{\nu}{2}}} y^{\frac{\nu-1}{2}} e^{-\frac{y}{2}\left(1+\frac{t^2}{\nu}\right)} & \text{para } y > 0 \text{ y } -\infty \to t \to \infty \\ 0 & \text{en cualquier otra parte} \end{cases}$$

y, al eliminar y por integración con la ayuda de la sustitución $w = \frac{y}{2}\left(1 + \frac{t^2}{\nu}\right)$, finalmente obtenemos

$$f(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\,\Gamma\left(\frac{\nu}{2}\right)} \cdot \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} \qquad \text{para } -\infty < t < \infty \qquad \blacktriangledown$$

Originalmente W. S. Gosset introdujo la distribución t, y publicaba sus escritos científicos bajo el seudónimo "Student", puesto que la compañía para la que trabajaba, una cervecería, no permitía que sus empleados hicieran publicaciones. Así, la distribución t también se conoce como la **distribución Student-t**, o la **distribución t de Student**.

En vista de su importancia, la distribución t se ha tabulado extensamente. Por ejemplo, la tabla IV contiene los valores de $t_{\alpha,\nu}$ para $\alpha = 0.10$, 0.05, 0.025, 0.01 y 0.005, y $\nu = 1, 2, \ldots, 29$, donde $t_{\alpha,\nu}$ es tal que el área a su derecha bajo la curva de la distribución t con ν grados de libertad (véase la figura 8.2) es igual a α. Esto es, $t_{\alpha,\nu}$ es tal que si T es una variable aleatoria que tiene una distribución t con ν grados de libertad, entonces

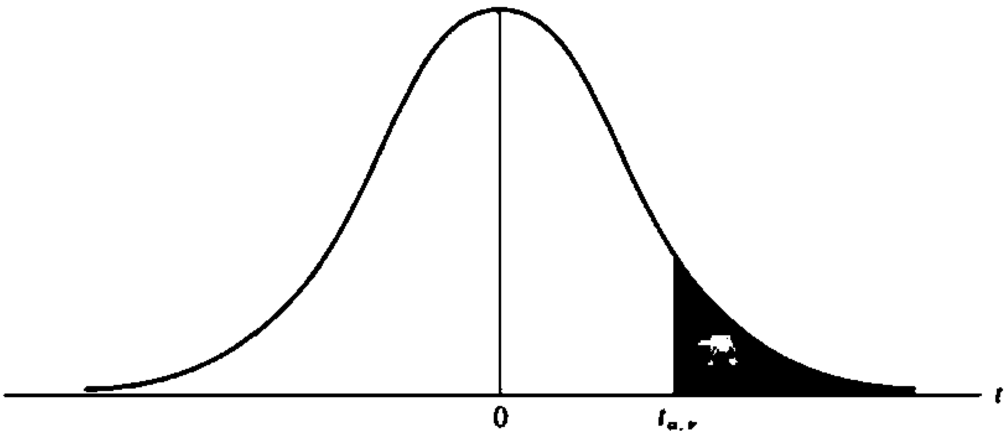

Figura 8.2 Distribución t.

$$P(T \geqq t_{\alpha,\nu}) = \alpha$$

La tabla no contiene valores de $t_{\alpha,\nu}$ para $\alpha > 0.50$, puesto que la densidad es simétrica con respecto a $t = 0$ y por tanto $t_{1-\alpha,\nu} = -t_{\alpha,\nu}$. Cuando ν es 30 o más, las probabilidades referentes a la distribución t suelen aproximarse con el uso de distribuciones normales (véase el ejercicio 8.50).

Entre tantas aplicaciones de la distribución t, algunas de las cuales se tratarán en los capítulos 11 y 13, su principal aplicación (para la cual se desarrolló originalmente) se basa en el siguiente teorema.

TEOREMA 8.13 Si $\overline{X}$ y S^2 son la media y la varianza de una muestra aleatoria de tamaño n de una población normal con la media μ y la varianza σ^2, entonces

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

tiene la distribución t con $n - 1$ grados de libertad.

Demostración. Por los teoremas 8.11 y 8.4, las variables aleatorias

$$Y = \frac{(n-1)S^2}{\sigma^2} \quad \text{y} \quad Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

tienen, respectivamente, una distribución ji cuadrada con $n - 1$ grados de libertad y la distribución normal estándar. Puesto que también son independientes por la parte 1 del teorema 8.11, la sustitución en la fórmula para T del teorema 8.12 nos da

$$T = \frac{\dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}}}{\sqrt{S^2/\sigma^2}} = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

y esto completa la demostración. ▼

EJEMPLO 8.3

En 16 corridas de prueba de una hora, el consumo de gasolina de una máquina promedió 16.4 galones con una desviación estándar de 2.1 galones. Pruebe la afirmación de que el consumo promedio de gasolina es de 12.0 galones por hora.

Solución

Al sustituir $n = 16$, $\mu = 12.0$, $\bar{x} = 16.4$, y $s = 2.1$ en la fórmula para t en el teorema 8.13, obtenemos

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{16.4 - 12.0}{2.1/\sqrt{16}} = 8.38$$

Puesto que la tabla IV muestra que para $\nu = 15$ la probabilidad de obtener un valor de T mayor que 2.947 es 0.005, la probabilidad de obtener un valor mayor que 8 debe ser despreciable. Así, parecería razonable concluir que el verdadero promedio de consumo de gasolina por hora de la máquina excede a 12.0 galones. ▲

8.6 LA DISTRIBUCIÓN *F*

Otra distribución con un papel importante en relación con el muestreo de poblaciones normales de la distribución F, llamada así en honor a Sir Ronald A. Fisher, uno de los estadísticos más prominentes de este siglo. Originalmente, se estudió como la distribución de muestreo de la razón de dos variables aleatorias independientes con distribuciones ji cuadrada, cada una dividida por sus grados de libertad respectivos, y es así como la presentaremos aquí.

TEOREMA 8.14 Si U y V son variables aleatorias independientes que tienen distribuciones independientes con ν_1 y ν_2 grados de libertad, entonces

$$F = \frac{U/\nu_1}{V/\nu_2}$$

es una variable aleatoria que tiene la **distribución *F***, esto es, una variable aleatoria cuya densidad de probabilidad está dada por

$$g(f) = \frac{\Gamma\left(\frac{\nu_1 + \nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)}\left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} \cdot f^{\frac{\nu_1}{2}-1}\left(1 + \frac{\nu_1}{\nu_2}f\right)^{-\frac{1}{2}(\nu_1+\nu_2)}$$

para $f > 0$ y $g(f) = 0$ en cualquier otra parte.

Demostración. La densidad conjunta de U y V está dada por

$$f(u, v) = \frac{1}{2^{\nu_1/2}\Gamma\left(\frac{\nu_1}{2}\right)} \cdot u^{\frac{\nu_1}{2}-1} e^{-\frac{u}{2}} \cdot \frac{1}{2^{\nu_2/2}\Gamma\left(\frac{\nu_2}{2}\right)} \cdot v^{\frac{\nu_2}{2}-1} e^{-\frac{v}{2}}$$

$$= \frac{1}{2^{(\nu_1+\nu_2)/2}\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} \cdot u^{\frac{\nu_1}{2}-1} v^{\frac{\nu_2}{2}-1} e^{-\frac{u+v}{2}}$$

para $u > 0$ y $v > 0$, y $f(u, v) = 0$ en cualquier otra parte. Entonces, para usar la técnica del cambio de variable de la sección 7.3, despejamos

$$f = \frac{u/v_1}{v/v_2}$$

para u, y obtenemos $u = \frac{\nu_1}{\nu_2} \cdot vf$ y por tanto $\frac{\partial u}{\partial f} = \frac{\nu_1}{\nu_2} \cdot v$. Así, por el teorema 7.1, la densidad conjunta de F y V está dada por

$$g(f, v) = \frac{\left(\frac{\nu_1}{\nu_2}\right)^{\nu_1/2}}{2^{(\nu_1+\nu_2)/2}\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} \cdot f^{\frac{\nu_1}{2}-1} v^{\frac{\nu_1+\nu_2}{2}-1} e^{-\frac{v}{2}\left(\frac{\nu_1 f}{\nu_2}+1\right)}$$

para $f > 0$ y $v > 0$, y $g(f, v) = 0$ en cualquier otra parte. Ahora, eliminamos v por integración al hacer la sustitución $w = \frac{v}{2}\left(\frac{\nu_1 f}{\nu_2} + 1\right)$, y finalmente obtenemos

$$g(f) = \frac{\Gamma\left(\frac{\nu_1 + \nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} \left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} \cdot f^{\frac{\nu_1}{2}-1}\left(1 + \frac{\nu_1}{\nu_2}f\right)^{-\frac{1}{2}(\nu_1+\nu_2)}$$

para $f > 0$, y $g(f) = 0$ en cualquier otra parte. ▼

En vista de su importancia, la distribución F se ha tabulado extensamente. Por ejemplo, la tabla VI contiene valores de f_{α,ν_1,ν_2} para $\alpha = 0.05$ y 0.01, y para diversos valores de ν_1 y ν_2, donde f_{α,ν_1,ν_2} es tal que el área a su derecha bajo la curva de la distribución F con ν_1 y ν_2 grados de libertad (véase la figura 8.3) es igual a α. Esto es f_{α,ν_1,ν_2} es tal que

$$P(F \geqq f_{\alpha,\nu_1,\nu_2}) = \alpha$$

Las aplicaciones del teorema 8.14 se presentan en problemas en los que nos interesa comparar las varianzas σ_1^2 y σ_2^2 de dos poblaciones normales; por ejemplo, en problemas en los que queremos estimar la razón $\frac{\sigma_1^2}{\sigma_2^2}$ o quizá probar si $\sigma_1^2 = \sigma_2^2$. Basamos

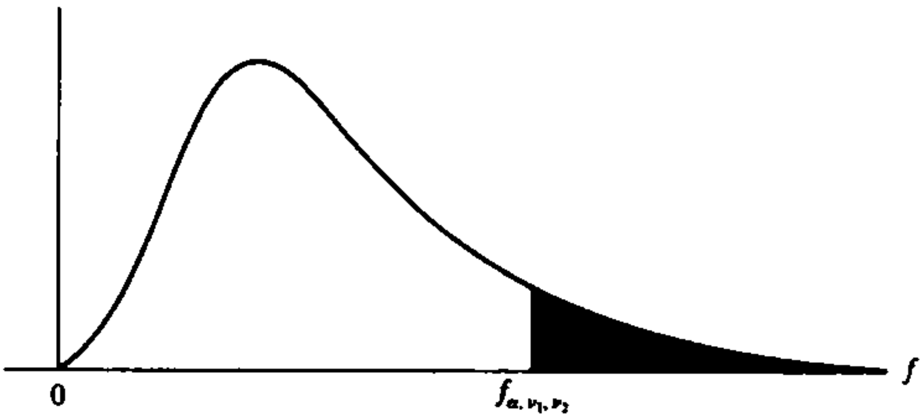

Figura 8.3 Distribución F.

tales inferencias en **muestras aleatorias independientes** de tamaño n_1 y n_2 de dos poblaciones y en el teorema 8.11, de acuerdo al cual

$$\chi_1^2 = \frac{(n_1 - 1)s_1^2}{\sigma_1^2} \quad \text{y} \quad \chi_2^2 = \frac{(n_2 - 1)s_2^2}{\sigma_2^2}$$

son los valores de variables aleatorias que tienen distribuciones ji cuadrada con $n_1 - 1$ y $n_2 - 1$ grados de libertad. Por "variables aleatorias independientes" queremos decir que las $n_1 + n_2$ variables aleatorias que constituyen las dos muestras aleatorias son todas independientes, de manera que las dos variables aleatorias ji cuadrada son independientes y la sustitución de sus valores por U y V en el teorema 8.14 produce el siguiente resultado.

TEOREMA 8.15 Si S_1^2 y S_2^2 son las varianzas de las muestras aleatorias independientes de tamaño n_1 y n_2 de poblaciones normales con las varianzas σ_1^2 y σ_2^2, entonces

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2}$$

es una variable aleatoria que tiene la distribución F con $n_1 - 1$ y $n_2 - 1$ grados de libertad.

En el capítulo 11 aplicaremos este teorema al problema de estimar la razón $\frac{\sigma_1^2}{\sigma_2^2}$ cuando se desconocen las varianzas de estas dos poblaciones; también, en el capítulo 13 demostraremos cómo probar si $\sigma_1^2 = \sigma_2^2$. En los procedimientos del análisis de la varianza del capítulo 15 se presentan aún otras pruebas basadas en la distribución F. Puesto que todas estas aplicaciones se basan en las razones de varianzas de muestras, la distribución F también se conoce como la **distribución de la razón de varianzas**.

EJERCICIOS

8.35 Demuestre el teorema 8.9.

8.36 Demuestre el teorema 8.10.

8.37 Verifique la identidad

$$\sum_{i=1}^{n}(X_i - \mu)^2 = \sum_{i=1}^{n}(X_i - \overline{X})^2 + n(\overline{X} - \mu)^2$$

que usamos en la demostración del teorema 8.11.

8.38 Use el teorema 8.11 para mostrar que, para muestras aleatorias de tamaño n de una población normal con la varianza σ^2, la distribución de muestreo S^2 tiene la media σ^2 y la varianza $\dfrac{2\sigma^4}{n-1}$. (En el libro de H. Cramér, incluido entre las referencias al final de este capítulo, se puede encontrar una fórmula general para la varianza de S^2 de muestras aleatorias de cualquier población con segundo y cuarto momentos finitos.)

8.39 Demuestre que si $X_1, X_2, \ldots, X_n$ son variables aleatorias independientes que tienen la distribución ji cuadrada con $\nu = 1$ y $Y_n = X_1 + X_2 + \cdots + X_n$, entonces la distribución límite de

$$Z = \frac{\dfrac{Y_n}{n} - 1}{\sqrt{2/n}}$$

conforme $n \to \infty$ es la distribución normal estándar.

8.40 Con base en los resultados del ejercicio 8.39, muestre que si X es una variable aleatoria que tiene una distribución ji cuadrada con ν grados de libertad y ν es grande, la distribución de $\dfrac{X-\nu}{\sqrt{2\nu}}$ se puede aproximar con la distribución normal estándar.

8.41 Use el método del ejercicio 8.40 para encontrar el valor aproximado de la probabilidad de que una variable aleatoria que tiene una distribución ji cuadrada con $\nu = 50$ asumirá un valor mayor que 68.0.

8.42 Si la amplitud de X es un conjunto de todos los números reales positivos, muestre que para $k > 0$ la probabilidad de que $\sqrt{2X} - \sqrt{2\nu}$ asumirá un valor menor que k es igual a la probabilidad de que $\dfrac{X-\nu}{\sqrt{2\nu}}$ asumirá un valor menor que $k + \dfrac{k^2}{2\sqrt{2\nu}}$.

8.43 Use los resultados de los ejercicios 8.40 y 8.42 para mostrar que si X tiene una distribución ji cuadrada con ν grados de libertad, entonces para ν grande la distribución de $\sqrt{2X} - \sqrt{2\nu}$ se puede aproximar con la distribución normal estándar. También, use este método de aproximación para volver a resolver el ejercicio 8.41.

8.44 Encuentre el porcentaje de error de las aproximaciones del ejercicio 8.41 y 8.43, dado que el valor real de la probabilidad (redondeado a cinco decimales) es 0.04596.

8.45 (*Demostración de la independencia de $\overline{X}$ y S^2 para $n = 2$.*) Si X_1 y X_2 son variables aleatorias independientes que tienen la distribución normal estándar, muestre que

(a) la densidad conjunta de X_1 y $\overline{X}$ está dada por

$$f(x_1, \bar{x}) = \frac{1}{\pi} \cdot e^{-\bar{x}^2} e^{-(x_1 - \bar{x})^2}$$

para $-\infty < x_1 < \infty$ y $-\infty < \bar{x} < \infty$;

(b) la densidad conjunta de $U = |X_1 - \overline{X}|$ y $\overline{X}$ está dada por

$$g(u, \bar{x}) = \frac{2}{\pi} \cdot e^{-(\bar{x}^2 + u^2)}$$

para $u > 0$ y $-\infty < \bar{x} < \infty$, puesto que $f(x_1, \bar{x})$ es simétrica alrededor de $\bar{x}$ para $\bar{x}$; fija;

(c) $S^2 = 2(X_1 - \overline{X})^2 = 2U^2$;

(d) la densidad conjunta de $\overline{X}$ y S^2 está dada por

$$h(s^2, \bar{x}) = \frac{1}{\sqrt{\pi}} e^{-\bar{x}^2} \cdot \frac{1}{\sqrt{2\pi}} (s^2)^{-\frac{1}{2}} e^{-\frac{1}{2}s^2}$$

para $s^2 > 0$ y $-\infty < \bar{x} < \infty$, de manera que $\overline{X}$ y S^2 son independientes.

8.46 (*Demostración de la independencia de $\overline{X}$ y S^2.*) Si $X_1, X_2, \ldots, X_n$ constituye una muestra aleatoria de una población normal con la media μ y la varianza σ^2,

(a) encuentre la densidad condicional de X_1 dado $X_2 = x_2$, $X_3 = x_3, \ldots, X_n = x_n$, y luego haga $X_1 = n\overline{X} - X_2 - \cdots - X_n$ y use la técnica de la transformación para encontrar la densidad condicional de $\overline{X}$ dado $X_2 = x_2$, $X_3 = x_3, \ldots, X_n = x_n$;

(b) encuentre la densidad conjunta de $\overline{X}, X_2, X_3, \ldots, X_n$ al multiplicar la densidad condicional de $\overline{X}$ obtenida en el inciso (a) por la densidad conjunta de $X_2, X_3, \ldots, X_n$, y muestre que

$$g(x_2, x_3, \ldots, x_n | \bar{x}) = \sqrt{n} \left(\frac{1}{\sigma\sqrt{2\pi}} \right)^{n-1} e^{-\frac{(n-1)s^2}{2\sigma^2}}$$

para $-\infty < x_i < \infty$, $i = 2, 3, \ldots, n$;

(c) muestre que la función generatriz de momentos condicional de $\dfrac{(n-1)S^2}{\sigma^2}$ dado $\overline{X} = \bar{x}$ es

$$E\left[e^{\frac{(n-1)S^2}{\sigma^2} \cdot t} \middle| \bar{x} \right] = (1 - 2t)^{-\frac{n-1}{2}} \qquad \text{para } t < \frac{1}{2}$$

Puesto que este resultado no tiene $\bar{x}$, se sigue que $\overline{X}$ y S^2 son independientes; también muestra que $\dfrac{(n-1)S^2}{\sigma^2}$ tiene la distribución ji cuadrada con $n - 1$ grados de libertad.

Esta demostración, que se debe a J. Shuster, se encuentra entre las referencias al final del capítulo.

8.47 Use la técnica de la transformación basada en el teorema 7.2 para rehacer la demostración del teorema 8.12. (*Sugerencia*: sean $t = \frac{z}{\sqrt{y/\nu}}$ y $u = y$.)

8.48 Muestre que para $\nu > 2$ la varianza de la distribución t con ν grados de libertad es $\frac{\nu}{\nu - 2}$. (*Sugerencia*: haga la sustitución $1 + \frac{t^2}{\nu} = \frac{1}{u}$.)

8.49 Muestre que para la distribución t con $\nu > 4$ grados de libertad

(a) $\mu_4 = \frac{3\nu^2}{(\nu - 2)(\nu - 4)}$;

(b) $\alpha_4 = 3 + \frac{6}{\nu - 4}$.

(*Sugerencia*: haga la sustitución $1 + \frac{t^2}{\nu} = \frac{1}{u}$.)

8.50 Use la fórmula de Stirling del ejercicio 1.6 para mostrar que cuando $\nu \to \infty$, la distribución t se aproxima a la distribución normal estándar.

8.51 ¿Con qué nombre nos referimos a la distribución t con $\nu = 1$ grado de libertad?

8.52 Use la técnica de la transformación basada en el teorema 7.2 para rehacer la demostración del teorema 8.14. (*Sugerencia*: sean $f = \frac{u/\nu_1}{v/\nu_2}$ y $w = v$.)

8.53 Muestre que para $\nu_2 > 2$ la media de la distribución F es $\frac{\nu_2}{\nu_2 - 2}$, haga uso de la definición de F en el teorema 8.14 y del hecho de que para una variable aleatoria V que tiene la distribución ji cuadrada con ν_2 grados de libertad, $E\left(\frac{1}{V}\right) = \frac{1}{\nu_2 - 2}$.

8.54 Verifique que si X tiene una distribución F con ν_1 y ν_2 grados de libertad y $\nu_2 \to \infty$, la distribución de $Y = \nu_1 X$ se aproxima a la distribución ji cuadrada con ν_1 grados de libertad.

8.55 Verifique que si T tiene una distribución t con ν grados de libertad, entonces $X = T^2$ tiene la distribución F con $\nu_1 = 1$ y $\nu_2 = \nu$ grados de libertad

8.56 Si X tiene una distribución F con ν_1 y ν_2 grados de libertad, muestre que $Y = \frac{1}{X}$ tiene la distribución F con ν_2 y ν_1 grados de libertad.

8.57 Use el resultado del ejercicio 8.56 para mostrar que

$$f_{1-\alpha, \nu_1, \nu_2} = \frac{1}{f_{\alpha, \nu_2, \nu_1}}$$

8.58 Verifique que si Y tiene la distribución beta con $\alpha = \frac{\nu_1}{2}$ y $\beta = \frac{\nu_2}{2}$, entonces

$$X = \frac{\nu_2 Y}{\nu_1(1 - Y)}$$

tiene la distribución F con ν_1 y ν_2 grados de libertad.

8.59 Muestre que la distribución F con 4 y 4 grados de libertad está dada por

$$g(f) = \begin{cases} 6f(1 + f)^{-4} & \text{para } f > 0 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

y use esta densidad para encontrar la probabilidad de que para las muestras aleatorias independientes de tamaño $n = 5$ de poblaciones normales con la misma varianza, S_1^2/S_2^2 asumirá un valor menor que $\frac{1}{2}$ o mayor que 2.

APLICACIONES

(En los ejercicios 8.61 hasta 8.66, consulte las tablas IV, V y VI.)

8.60 Integre la densidad ji cuadrada apropiada para encontrar la probabilidad de que la varianza de una muestra aleatoria de tamaño 5 de una población normal con $\sigma^2 = 25$ caerá entre 20 y 30.

8.61 La afirmación de que la varianza de una población normal es $\sigma^2 = 25$ debe rechazarse si la varianza de una muestra aleatoria de tamaño 16 excede 54.668 o es menor a 12.102. ¿Cuál es la probabilidad de que esta afirmación se rechazará aun cuando $\sigma^2 = 25$?

8.62 La afirmación de que la varianza de una población normal es $\sigma^2 = 4$ debe rechazarse si la varianza de una muestra aleatoria de tamaño 9 excede 7.7535. ¿Cuál es la probabilidad de que esta afirmación se rechazará aun cuando $\sigma^2 = 4$?

8.63 Una muestra aleatoria de tamaño $n = 25$ de una población normal que tiene la media $\bar{x} = 47$ y la desviación estándar $s = 7$. Si basamos nuestra decisión en la estadística del teorema 8.13, ¿podemos decir que la información dada sustenta la conjetura de que la media de la población es $\mu = 42$?

8.64 Una muestra aleatoria de tamaño $n = 12$ de una población normal tiene la media $\bar{x} = 27.8$ y la varianza $s^2 = 3.24$. Si basamos nuestra decisión en la estadística del teorema 8.13, ¿podemos decir que la información dada sustenta la afirmación de que la media de la población es $\mu = 28.5$?

8.65 Si S_1 y S_2 son las desviaciones estándar de variables aleatorias independientes de tamaños $n_1 = 61$ y $n_2 = 31$ de poblaciones normales con $\sigma_1^2 = 12$ y $\sigma_2^2 = 18$, encuentre $P(S_1^2/S_2^2 > 1.16)$.

8.66 Si S_1^2 y S_2^2 son las varianzas de las variables aleatorias independientes de tamaños $n_1 = 10$ y $n_2 = 15$ de poblaciones normales con varianzas iguales, encuentre $P(S_1^2/S_2^2 < 4.03)$.

8.67 Use un programa de computadora para verificar los cinco elementos de la tabla IV que corresponden a 11 grados de libertad.

8.68 Use un programa de computadora para verificar los ocho elementos de la tabla V que corresponden a 21 grados de libertad.

8.69 Use un programa de computadora para verificar los cinco valores de $f_{0.05}$ en la tabla VI que corresponden a 7 y 6 a 10 grados de libertad.

8.70 Use un programa de computadora para verificar los seis valores de $f_{0.01}$ en la tabla VI que corresponden a 12 a 17 y 16 grados de libertad.

8.7 ESTADÍSTICAS DE ORDEN

Considere una muestra aleatoria de tamaño n de una población infinita que tiene una densidad continua, y suponga que arreglamos los valores de $X_1, X_2, \ldots,$ y X_n de acuerdo a su tamaño. Si consideramos la más pequeña de las x como un valor de la variable aleatoria Y_1, el siguiente valor en tamaño como un valor de la variable aleatoria Y_2, el siguiente valor en tamaño como un valor de la variable aleatoria $Y_3, \ldots,$ y el valor más grande como un valor de la variable aleatoria Y_n, nos referimos a estas Y como **estadísticas de orden**. En particular, Y_1 es la estadística de primer orden, Y_2 es la estadística de segundo orden, Y_3 es la estadística de tercer orden, y así sucesivamente. (Estamos limitando este examen a poblaciones infinitas con densidades continuas así que hay una probabilidad de cero que dos valores cualquiera de las x serán iguales.)

Para ser más explícitos, considere el caso donde $n = 2$ y la relación entre los valores del las X y las Y es

$$y_1 = x_1 \quad \text{y} \quad y_2 = x_2 \quad \text{cuando} \quad x_1 < x_2$$

$$y_1 = x_2 \quad \text{y} \quad y_2 = x_1 \quad \text{cuando} \quad x_2 < x_1$$

De manera similar, para $n = 3$ la relación entre los valores de las variables aleatorias respectivas es

$$y_1 = x_1, \quad y_2 = x_2, \quad \text{y} \quad y_3 = x_3, \quad \text{cuando} \quad x_1 < x_2 < x_3$$

$$y_1 = x_1, \quad y_2 = x_3, \quad \text{y} \quad y_3 = x_2, \quad \text{cuando} \quad x_1 < x_3 < x_2$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$y_1 = x_3, \quad y_2 = x_2, \quad \text{y} \quad y_3 = x_1, \quad \text{cuando} \quad x_3 < x_2 < x_1$$

Derivemos ahora una fórmula para la densidad de probabilidad de la $r$ésima estadística de orden para $r = 1, 2, \ldots, n$.

TEOREMA 8.16 Para muestras aleatorias de tamaño n de una población infinita que tiene el valor $f(x)$ en x, la densidad de probabilidad de la $r$ésima estadística de orden Y_r está dada por

$$g_r(y_r) = \frac{n!}{(r-1)!(n-r)!}\left[\int_{-\infty}^{y_r} f(x)\,dx\right]^{r-1} f(y_r)\left[\int_{y_r}^{\infty} f(x)\,dx\right]^{n-r}$$

para $-\infty < y_r < \infty$.

Demostración. Supongamos que el eje real está dividido en tres intervalos, uno de $-\infty$ a y_r, un segundo de y_r a $y_r + h$ (donde h es una constante positiva), y el tercero de $y_r + h$ a ∞. Puesto que la población que estamos muestreando tiene el valor $f(x)$ en x, la probabilidad de que $r - 1$ de los valores de la muestra caigan en el primer intervalo, uno caiga en el segundo intervalo, y $n - r$ caigan en el tercer intervalo es

$$\frac{n!}{(r-1)!1!(n-r)!}\left[\int_{-\infty}^{y_r} f(x)\,dx\right]^{r-1}\left[\int_{y_r}^{y_r+h} f(x)\,dx\right]\left[\int_{y_r+h}^{\infty} f(x)\,dx\right]^{n-r}$$

de acuerdo a la fórmula para la distribución multinomial. Usamos la ley de la media del cálculo, y tenemos

$$\int_{y_r}^{y_r+h} f(x)\,dx = f(\xi)\cdot h \qquad \text{cuando } y_r \leqq \xi \leqq y_r + h$$

y si hacemos $h \to 0$, obtenemos finalmente

$$g_r(y_r) = \frac{n!}{(r-1)!(n-r)!}\left[\int_{-\infty}^{y_r} f(x)\,dx\right]^{r-1} f(y_r)\left[\int_{y_r}^{\infty} f(x)\,dx\right]^{n-r}$$

para $-\infty < y_r < \infty$ para la densidad de probabilidad de la $r$ésima estadística de orden. ▼

En particular, la distribución de muestreo de Y_1, el valor más pequeño en la muestra aleatoria de tamaño n, está dada por

$$g_1(y_1) = n\cdot f(y_1)\left[\int_{y_1}^{\infty} f(x)\,dx\right]^{n-1} \qquad \text{para } -\infty < y_1 < \infty$$

mientras que la distribución de muestreo de Y_n, el valor más grande en una muestra aleatoria de tamaño n, está dada por

$$g_n(y_n) = n\cdot f(y_n)\left[\int_{-\infty}^{y_n} f(x)\,dx\right]^{n-1} \qquad \text{para } -\infty < y_n < \infty$$

También, en una muestra aleatoria de tamaño $n = 2m + 1$ la **mediana de la muestra** $\widetilde{X}$ es Y_{m+1}, cuya distribución de muestreo está dada por

$$h(\tilde{x}) = \frac{(2m+1)!}{m!m!}\left[\int_{-\infty}^{\tilde{x}} f(x)\,dx\right]^{m} f(\tilde{x})\left[\int_{\tilde{x}}^{\infty} f(x)\,dx\right]^{m} \qquad \text{para } -\infty < \tilde{x} < \infty$$

[Para muestras aleatorias de tamaño $n = 2m$, la mediana se define como $\frac{1}{2}(Y_m + Y_{m+1})$.]

En algunos casos es posible efectuar las integraciones requeridas para obtener las densidades de las diversas estadísticas de orden; para otras poblaciones tal vez no haya otra opción que aproximar estas integrales usando métodos numéricos.

EJEMPLO 8.4

Muestre que para una muestra aleatoria de tamaño n de una población exponencial con el parámetro θ, las distribuciones de muestreo de Y_1 y Y_n están dadas por:

$$g_1(y_1) = \begin{cases} \dfrac{n}{\theta} \cdot e^{-ny_1/\theta} & \text{para } y_1 > 0 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

y

$$g_n(y_n) = \begin{cases} \dfrac{n}{\theta} \cdot e^{-y_n/\theta}[1 - e^{-y_n/\theta}]^{n-1} & \text{para } y_n > 0 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

y que, para muestras aleatorias de tamaño $n = 2m + 1$ de esta clase de población, la distribución de muestreo de la mediana está dada por

$$h(\tilde{x}) = \begin{cases} \dfrac{(2m + 1)!}{m!m!\theta} \cdot e^{-\tilde{x}(m+1)/\theta}[1 - e^{-\tilde{x}/\theta}]^m & \text{para } \tilde{x} > 0 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

Solución

Las integraciones requeridas para obtener estos resultados son simples, y se dejarán al lector en el ejercicio 8.71. ▲

El siguiente es un resultado interesante sobre la distribución de muestreo de la mediana, que es válido cuando la densidad de población es continua y distinta de cero en la **mediana de la población** $\tilde{\mu}$, que es tal que $\int_{-\infty}^{\tilde{\mu}} f(x)\, dx = \frac{1}{2}$.

TEOREMA 8.17 Para n grande, la distribución de muestreo de la mediana para muestras aleatorias de tamaño $2n + 1$ es aproximadamente normal con la media $\tilde{\mu}$ y la varianza $\dfrac{1}{8[f(\tilde{\mu})]^2 n}$.

En la página 298 se hace referencia a una demostración de este teorema. Advierta que para muestras aleatorias de tamaño $2n + 1$ de una población normal tenemos $\mu = \tilde{\mu}$, así que

$$f(\tilde{\mu}) = f(\mu) = \frac{1}{\sigma\sqrt{2\pi}}$$

y la varianza de la mediana es aproximadamente $\dfrac{\pi\sigma^2}{4n}$. Si comparamos esto con la varianza de la media, que para muestras aleatorias de tamaño $2n + 1$ de una población infinita es $\dfrac{\sigma^2}{2n + 1}$, encontramos que para muestras grandes de poblaciones normales la media es **más confiable** que la mediana; esto es, la media está sujeta a fluctuaciones fortuitas menores que la mediana.

EJERCICIOS

8.71 Verifique los resultados del ejemplo 8.4, esto es, las distribuciones de muestreo de Y_1, Y_n y $\tilde{X}$ mostrados ahí para muestras aleatorias de una población exponencial.

8.72 Encuentre las distribuciones de muestreo de Y_1 y Y_n para las muestras aleatorias de tamaño n de una población uniforme continua con $\alpha = 0$ y $\beta = 1$.

8.73 Encuentre la distribución de muestreo de la mediana para muestras aleatorias de tamaño $2m + 1$ de la población del ejercicio 8.72.

8.74 Encuentre la media y la varianza de la distribución de muestreo de Y_1 para muestras aleatorias de tamaño n de la población del ejercicio 8.72.

8.75 Encuentre las distribuciones de muestreo de Y_1 y Y_n para muestras aleatorias de tamaño n de una población que tiene la distribución beta con $\alpha = 3$ y $\beta = 2$.

8.76 Encuentre la distribución de muestreo de la mediana para muestras aleatorias de tamaño $2m + 1$ de la población del ejercicio 8.75.

8.77 Encuentre la distribución de muestreo de Y_1 para muestras aleatorias de tamaño $n = 2$ tomadas

(a) sin reemplazo de la población finita que consiste en los primeros cinco enteros positivos;

(b) con reemplazo de la misma población.

(*Sugerencia*: enumere todas las posibilidades.)

8.78 Repita el método usado en la demostración del teorema 8.16 para demostrar que la densidad conjunta de Y_1 y Y_n está dada por

$$g(y_1, y_n) = n(n-1)f(y_1)f(y_n)\left[\int_{y_1}^{y_n} f(x)\,dx\right]^{n-2} \quad \text{para } -\infty \rightarrow y_1 \rightarrow y_n \rightarrow \infty$$

y $g(y_1, y_n) = 0$ en cualquier otra parte.

(a) Use este resultado para encontrar la densidad conjunta de Y_1 y Y_n para muestras aleatorias de tamaño n de una población exponencial.

(b) Use este resultado para encontrar la densidad conjunta de Y_1 y Y_n para la población del ejercicio 8.72.

8.79 Con respecto al inciso (b) del ejercicio 8.78, encuentre la covarianza de Y_1 y Y_n.

8.80 Use la fórmula de la densidad conjunta de Y_1 y Y_n que se muestra en el ejercicio 8.78 y la técnica de transformación de la sección 7.4 para encontrar una expresión para la densidad conjunta de Y_1 y la **amplitud de la muestra** $R = Y_n - Y_1$.

8.81 Use el resultado del ejercicio 8.80 y el del inciso (a) del ejercicio 8.78 para encontrar la distribución de muestreo de R para muestras aleatorias de tamaño n de una población exponencial.

8.82 Use el resultado del ejercicio 8.80 para encontrar la distribución de muestreo de R para muestras aleatorias de tamaño n de poblaciones exponenciales de la población uniforme continua del ejercicio 8.72.

8.83 Use el resultado del ejercicio 8.82 para encontrar la media y la varianza de la distribución de muestreo de R para muestras aleatorias de tamaño n de la población uniforme continua del ejercicio 8.72.

8.84 Hay muchos problemas, particularmente en aplicaciones industriales, donde nos interesa la proporción de una población que se encuentra entre ciertos límites. Tales límites se llaman **límites de tolerancia.** Los pasos siguientes nos llevan a la distribución de muestreo de la estadística P, que es la proporción de una población (que tiene una densidad continua) que se encuentra entre los valores más pequeño y más grande de una muestra aleatoria de tamaño n.

(a) Use la fórmula para la densidad conjunta de Y_1 y Y_n que se muestra en el ejercicio 8.78 y la técnica de la transformación de la sección 7.4 para demostrar que la densidad conjunta de Y_1 y P, cuyos valores están dados por

$$p = \int_{y_1}^{y_n} f(x)\,dx$$

es

$$h(y_1, p) = n(n-1)f(y_1)p^{n-2}$$

(b) Use el resultado del inciso (a) y la técnica de la transformación de la sección 7.4 para demostrar que la densidad conjunta de P y W, cuyos valores están dados por

$$w = \int_{-\infty}^{y_1} f(x)\,dx$$

es

$$\varphi(w, p) = n(n-1)p^{n-2}$$

para $w > 0$, $p > 0$, $w + p \to 1$ y $\varphi(w, p) = 0$ en cualquier otra parte.

(c) Use el resultado del inciso (b) para demostrar que la densidad marginal de P está dada por

$$g(p) = \begin{cases} n(n-1)p^{n-2}(1-p) & \text{para } 0 < p < 1 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

Ésta es la densidad deseada de la proporción de población que se encuentra en el valor más pequeño y en el más grande de una muestra aleatoria de tamaño n, y resulta interesante observar que no depende de la forma de la distribución de la población.

8.85 Use el resultado del ejercicio 8.84 para demostrar que, para la variable aleatoria P definida ahí,

$$E(P) = \frac{n-1}{n+1} \quad \text{y} \quad \text{var}(P) = \frac{2(n-1)}{(n+1)^2(n+2)}$$

¿Qué podemos concluir de esto sobre la distribución de P cuando n es grande?

APLICACIONES

8.86 Encuentre la probabilidad de que en una muestra aleatoria de tamaño $n = 4$ de la población uniforme continua del ejercicio 8.72, el valor más pequeño será al menos 0.20.

8.87 Encuentre la probabilidad de que en una muestra aleatoria de tamaño $n = 3$ de la población beta del ejercicio 8.75, el valor más grande será al menos 0.90.

8.88 Use el resultado del ejercicio 8.82 para encontrar la probabilidad de que la amplitud de una muestra aleatoria de tamaño $n = 5$ de la población uniforme dada será al menos 0.75.

8.89 Use el resultado del inciso (c) del ejercicio 8.84 para encontrar la probabilidad de que en una muestra aleatoria de tamaño $n = 10$ al menos 80 por ciento de la población estará entre los valores más grandes y más pequeños.

8.90 Use el resultado del inciso (c) del ejercicio 8.84 para establecer una ecuación en n cuya solución dará el tamaño de la muestra que es requerida para poder afirmar con probabilidad $1 - \alpha$ que la proporción de la población contenida entre los valores más pequeños y más grandes de la muestra es al menos p. Demuestre que para $p = 0.90$ y $\alpha = 0.05$ esta ecuación se puede escribir como

$$(0.90)^{n-1} = \frac{1}{2n + 18}$$

Esta clase de ecuación es difícil de resolver, pero se puede mostrar que una solución aproximada para n está dada por

$$\frac{1}{2} + \frac{1}{4} \cdot \frac{1 + p}{1 - p} \cdot \chi^2_{\alpha, 4}$$

donde $\chi^2_{\alpha, 4}$ se debe buscar en la tabla V. Use este método para encontrar una solución aproximada a la ecuación para $p = 0.90$ y $\alpha = 0.05$.

REFERENCIAS

Condiciones necesarias y suficientes para la forma más sólida del teorema del límite central para variables aleatorias independientes, las condiciones *Lindberg-Feller*, se dan en

FELLER, W., *An Introduction to Probability Theory and Its Applications*, Vol. I, 3rd ed. Nueva York: John Wiley & Sons, Inc., 1968,

así como en otros textos avanzados sobre teoría de probabilidad.

Tablas extensas de las distribuciones normal, ji cuadrada, F, y t se pueden encontrar en

PEARSON, E. S., and HARTLEY, H. O., *Biometrika Tables for Statisticians*, Vol. I. Nueva York: John Wiley & Sons, Inc., 1968.

Una fórmula general para la varianza de la distribución de muestreo del segundo momento de la muestra M_2 (el cual difiere de S^2 sólo en lo que dividimos entre n en vez de $n - 1$) se deriva en

CRAMÉR, H., *Mathematical Methods of Statistics*. Princeton, N.J.: Princeton University Press, 1950,

y una demostración del teorema 8.17 se da en

WILKS, S. S., *Mathematical Statistics*. Nueva York: John Wiley & Sons, Inc., 1962.

En muchos textos avanzados sobre estadística matemática se dan demostraciones de la independencia de $\overline{X}$ y S^2. Por ejemplo, una demostración basada en la funciones generatrices de momentos se puede encontrar en el libro mencionado de W. S. Wilks, y una demostración un poco más elemental, ilustrada para $n = 3$, se puede encontrar en

KEEPING, E. S., *Introduction to Statistical Inference*. Princeton, N.J.: D. Van Nostrand Co., Inc., 1962.

La demostración bosquejada en el ejercicio 8.46 se encuentra en

SHUSTER, J., "A Simple Method of Teaching the Independence of $\overline{X}$ and S^2," *The American Statistician*, Vol. 27, No. 1, 1973.

CAPÍTULO

9

Teoría de decisiones†

9.1 INTRODUCCIÓN

En el capítulo 4 introdujimos el concepto de una esperanza matemática para estudiar los valores esperados de variables aleatorias, en particular los momentos de sus distribuciones. En situaciones aplicadas, las esperanzas matemáticas se usan a menudo como una guía para escoger entre alternativas, esto es, en la toma de decisiones, ya que generalmente se considera racional seleccionar alternativas con las esperanzas matemáticas "más prometedoras", aquellas que maximizan las utilidades esperadas, minimizan las pérdidas esperadas, maximizan las ventas esperadas, minimizan los costos esperados y así sucesivamente.

Aunque este enfoque hacia la toma de decisiones tiene una apariencia de gran intuición, no carece de complicaciones ya que hay varios problemas donde resulta difícil, si no es que imposible, asignar valores numéricos a las consecuencias de nuestras acciones y a las probabilidades de todas las eventualidades.

EJEMPLO 9.1

Un fabricante de artículos de cuero debe decidir si expande la capacidad de su planta ahora o espera al menos otro año. Sus asesores le dicen que si se expande ahora y las condiciones económicas siguen siendo buenas, habrá una utilidad de $164,000 durante el siguiente año fiscal; si se expande ahora y hay una recesión, habrá una pérdida de $40,000; si se espera al menos otro año y las condiciones económicas siguen siendo bue-

† Aunque el material aquí presentado es fundamental para un entendimiento de las bases de la estadística, a menudo se omite en un primer curso de estadística matemática. Se puede omitir sin pérdida alguna en la continuidad.

nas, habrá una utilidad de $80,000; si se espera al menos otro año y hay una recesión, habrá una pequeña utilidad de $8,000. ¿Qué es lo que el fabricante debe decidir hacer si desea minimizar su pérdida esperada durante el siguiente año fiscal y si siente que la ventaja es 2 a 1 de que habrá una recesión?

Solución

Esquemáticamente, todos estos "resultados" se pueden representar en la tabla siguiente, donde los elementos son las pérdidas que corresponden a las diversas posibilidades y, por tanto, las ganancias se representan con números negativos:

	Expandirse ahora	*Retrasar la expansión*
Las condiciones económicas siguen siendo buenas	−164,000	−80,000
Hay una recesión	40,000	−8,000

Aquí estamos trabajando con pérdidas en vez de utilidades para hacer que este ejemplo se ajuste al esquema general que presentaremos en las secciones 9.2 y 9.3.

Puesto que las probabilidades de que las condiciones económicas sigan siendo buenas y de que habrá una recesión son, respectivamente, $\frac{1}{3}$ y $\frac{2}{3}$, la pérdida esperada del fabricante para el siguiente año fiscal es

$$-164{,}000\cdot\frac{1}{3} + 40{,}000\cdot\frac{2}{3} = -28{,}000$$

si expande la capacidad de su planta ahora, y

$$-80{,}000\cdot\frac{1}{3} + (-8{,}000)\cdot\frac{2}{3} = -32{,}000$$

si se espera otro año. Puesto que una utilidad esperada (pérdida esperada negativa) de $32,000 es preferible a una utilidad esperada (pérdida esperada negativa) de $28,000, se sigue que el fabricante debe retrasar la expansión de la capacidad de su planta. ▲

El resultado al que llegamos en este ejemplo supone que los valores dados en la tabla y también la ventaja de una recesión se evalúan en forma correcta. Como se pedirá al lector que muestre en los ejercicios 9.2 y 9.3, los cambios en estas cantidades pueden llevar fácilmente a resultados diferentes.

EJEMPLO 9.2

Con respecto al ejemplo 9.1, supongamos que el fabricante no tiene idea acerca de la ventaja de que habrá una recesión. ¿Qué debe hacer si él es un pesimista empedernido?

Solución

Al ser el fabricante de la clase de persona que siempre espera que suceda lo peor, podría argumentar que si expande la capacidad de su planta ahora podría perder

\$40,000, si retrasa la expansión habría una utilidad de al menos \$8,000 y, por tanto, que minimizará la pérdida máxima (o maximizará la utilidad mínima) si se espera al menos otro año. ▲

El criterio que se usó en este ejemplo se llama el **criterio minimax**, y es sólo uno de tantos criterios diferentes que se pueden usar en esta clase de situaciones. En el ejercicio 9.7 se hace referencia a un criterio así, con base en el optimismo en vez del pesimismo, y en el ejercicio 9.8 se hace referencia a otro que se basa en el temor de "salir perdiendo en un buen negocio".

9.2 TEORÍA DE JUEGOS

Los ejemplos de la sección anterior bien pueden haber dado la impresión que el fabricante está dentro de un juego, un juego entre él y la Naturaleza (o llámelo fortuna o lo "que controla" si hay o no una recesión). Cada uno de los "jugadores" tiene la opción de dos movimientos: el fabricante tiene la opción entre las acciones a_1 y a_2 (expandir la capacidad de su planta ahora o retrasar la expansión al menos un año), y la Naturaleza controla la elección entre θ_1 y θ_2 (si las condiciones económicas siguen siendo buenas o si habrá una recesión). Dependiendo de la elección de sus movimientos, se obtienen los "resultados" mostrados en la tabla siguiente:

		Jugador A (El fabricante) a_1	a_2
Jugador B	θ_1	$L(a_1, \theta_1)$	$L(a_2, \theta_1)$
(Naturaleza)	θ_2	$L(a_1, \theta_2)$	$L(a_2, \theta_2)$

Las cantidades $L(a_1, \theta_1)$, $L(a_2, \theta_1), \ldots$, se conocen como los valores de la **función de pérdida** que caracteriza al "juego" en particular; en otras palabras, $L(a_i, \theta_j)$ es la pérdida del jugador A (la cantidad que tiene que pagar al jugador B) cuando escoge la alternativa a_i y el jugador B escoge la alternativa θ_j. Aunque en realidad no es importante, supondremos que estas cantidades están en dólares. En la práctica real, también se pueden expresar en términos de cualesquier mercancías o servicios, en unidades de utilidad ("deseabilidad" o satisfacción), y aun en términos de vida o muerte (como en la ruleta rusa o en el proceder de una guerra).

En realidad la analogía que aquí hacemos no es excéntrica; el problema del ejemplo 9.1 es típico de la clase de situaciones tratadas en la **teoría de juegos**, una rama relativamente nueva de las matemáticas que ha estimulado un interés considerable en años recientes. Esta teoría no se limita a los juegos de salón, como su nombre podría sugerir, sino que se aplica a cualquier clase de situación competitiva y, como veremos, ha llevado a un enfoque unificado para resolver problemas de inferencia estadística.

Para introducir algunos de los conceptos básicos de la teoría de juegos, empecemos por explicar qué queremos decir por un **juego de dos personas de suma cero**. En este término, "dos personas" significa que hay dos jugadores (o más generalmente, dos grupos con intereses encontrados), y "suma cero" significa que lo que un jugador pierde el otro jugador lo gana. Así, en un juego de suma cero no hay "participación del casino", como en los juegos profesionales, y no se crea ni se destruye capital durante el curso del juego. Por supuesto, la teoría de juegos también incluye juegos que no son ni suma cero ni están limitados a dos jugadores, pero, como bien nos podemos imaginar, tales juegos son generalmente mucho más complicados. El ejercicio 9.19 es ejemplo de un juego que no es de suma cero.

Los juegos también se clasifican de acuerdo al número de **estrategias** (movimientos, opciones o alternativas) que cada jugador tiene a su disposición. Por ejemplo, si cada jugador tiene que escoger una de dos alternativas (como en el ejemplo 9.1) decimos que es un juego 2×2 si un jugador tiene 3 movimientos posibles mientras que el otro tiene 4, el juego es 3×4 o 4×3, según sea el caso. En esta sección sólo consideraremos juegos **finitos**, esto es, juegos donde cada jugador sólo tiene un número finito, o fijo, de movimientos posibles, pero más tarde también consideraremos juegos donde cada jugador tiene infinitamente muchos movimientos.

En la teoría de juegos se acostumbra referirse a los dos jugadores como jugador A y jugador B como hicimos en la tabla anterior, pero los movimientos (opciones o alternativas) del jugador A suelen etiquetarse como I, II, III, ..., en vez de $a_1, a_2, a_3, \ldots$, y las del jugador B suelen etiquetarse como 1, 2, 3, ..., en vez de $\theta_1, \theta_2, \theta_3, \ldots$. Los **resultados**, las cantidades del dinero que cambia de manos cuando los jugadores escogen sus estrategias respectivas, suelen mostrarse en una tabla como la de la página 302, la cual en la teoría de juegos se conoce como **matriz de resultados**. (Como antes, los resultados positivos representan pérdidas del jugador A y los resultados negativos representan pérdidas del jugador B). También mencionemos que en la teoría de juegos siempre se supone que cada jugador debe escoger su estrategia sin saber qué es lo que su oponente va a hacer y que una vez que un jugador hizo su elección no puede cambiarla.

Los objetivos de la teoría de juegos son determinar las **estrategias óptimas** (esto es, estrategias que son más redituables a los jugadores respectivos) y el resultado correspondiente, que se llama el **valor** del juego.

EJEMPLO 9.3

Dado el juego de dos personas de suma cero 2×2

		Jugador A	
		I	II
Jugador B	1	7	−4
	2	8	10

encuentre las estrategias óptimas de los jugadores A y B y el valor del juego.

Solución

Como se puede ver mediante una inspección, sería tonto para el jugador B escoger la estrategia 1, puesto que la estrategia 2 le rinde más que la estrategia 1 sin importar la elección que hizo el jugador A. En una situación como ésta decimos que la estrategia 1 **está dominada** por la estrategia 2 (o que la estrategia 2 **domina** a la estrategia 1) y es lógico que cualquier estrategia que está dominada por otra debe descartarse. Si aquí hacemos esto, encontramos que la estrategia óptima del jugador B es la estrategia 2, la única que queda, y que la estrategia óptima del jugador A es la estrategia I, puesto que una pérdida de \$8 es obviamente preferible a una pérdida de \$10. También el valor del juego, el resultado correspondiente a las estrategias I y 2, es \$8. ▲

EJEMPLO 9.4

Dado el juego de dos personas de suma cero 3×2

		Jugador A		
		I	II	III
Jugador B	1	−4	1	7
	2	4	3	5

encuentre las estrategias óptimas de los jugadores A y B y el valor del juego.

Solución

En este juego ninguna estrategia de la jugadora B domina a la otra, pero la tercera estrategia del jugador A está dominada por cada una de las otras dos; claramente, una utilidad de \$4 o una pérdida de \$1 es preferible a una pérdida de \$7, y una pérdida de \$4 o una pérdida de \$3 es preferible a una pérdida de \$5. Así, podemos descartar la tercera columna de la matriz de resultados y estudiar el juego 2×2

		Jugador A	
		I	II
Jugador B	1	−4	1
	2	4	3

donde ahora la estrategia 2 de la jugadora B domina a la estrategia 1. Así, la opción óptima para la jugadora B es la estrategia 2, la opción óptima del jugador A es la estrategia II (puesto que una pérdida de \$3 es preferible a una pérdida de \$4), y el valor del juego es \$3. ▲

El proceso de descartar las estrategias dominadas puede ser de gran ayuda en la solución de un juego (esto es, en encontrar las estrategias óptimas y el valor del juego), pero ésta es la excepción más que la regla que nos llevará a una solución completa. Es posible que no existieran dominancias, como se ilustra por el siguiente juego de dos personas de suma cero 3×3:

		Jugador A		
		I	II	III
	1	−1	6	−2
Jugador B	2	2	4	6
	3	−2	−6	12

Así, debemos buscar otros métodos para llegar a las estrategias óptimas. Desde el punto de vista del jugador A, podríamos argüir de la siguiente manera: si escoge la estrategia I, lo peor que le puede pasar es que pierda \$2; si escoge la estrategia II, lo peor que le puede pasar es que pierda \$6, y si escoge la estrategia III, lo peor que le puede pasar es que pierda \$12. Así, podría minimizar la pérdida máxima al escoger la estrategia I.

Aplicamos el mismo tipo de argumentación para seleccionar la estrategia para la jugadora B, encontramos que si escoge la estrategia 1, lo peor que le puede pasar es que pierda \$2, si escoge la estrategia 2, lo peor que le puede pasar es que gane \$2, y si escoge la estrategia 3, lo peor que le puede pasar es que pierda \$6. Así, puede minimizar su pérdida máxima (o maximizar su ganancia mínima, que es lo mismo) al escoger la estrategia 2.

La selección de las estrategias I y 2, apropiadamente llamadas estrategias **minimax** (o estrategias basadas en el **criterio minimax**), es realmente muy razonable. Al escoger la estrategia I, el jugador A se asegura que su oponente puede ganar cuando mucho \$2, y al escoger la estrategia 2, la jugadora B se asegura que realmente ganará esta cantidad. Estos \$2 son el valor del juego lo cual significa que el juego favorece a la jugadora B, pero lo podríamos hacer **equitativo** cobrándole a la jugadora B \$2 por el privilegio de estar en el juego y darle los \$2 al jugador A.

Un aspecto muy importante de las estrategias minimax I y 2 de este ejemplo es que son completamente a "prueba de espías" en el sentido que ningún jugador puede beneficiarse de conocer la elección del otro. En nuestro ejemplo, aun si el jugador A anunciara públicamente que escogerá la estrategia I, de todos modos sería mejor para el jugador B escoger la estrategia 2, y si el jugador B anunciara públicamente que escogerá la estrategia 2, aún sería mejor para el jugador A escoger la estrategia I. Desdichadamente, no todos los juegos son a prueba de espías.

EJEMPLO 9.5

Demuestre que las estrategias minimax de los jugadores A y B no son a prueba de espías en el siguiente juego

		Jugador A	
		I	II
Jugador B	1	8	−5
	2	2	6

Solución

El jugador A puede minimizar su pérdida máxima al escoger la estrategia II, y la jugadora B puede minimizar su pérdida máxima al escoger la estrategia 2. Sin embargo, si el jugador A supiera que la jugadora B va a basar su elección en el criterio minimax podría cambiar a la estrategia I y así reducir su pérdida de \$6 a \$2. Por supuesto, si la jugadora B descubriera que el jugador A tratará de ser más sagaz que ella en esta forma, podría a su vez cambiar a la estrategia 1 y aumentar su ganancia a \$8. En cualquier caso, las estrategias minimax de los dos jugadores no son a prueba de espías, lo que deja espacio para toda clase de trucos y engaños. ▲

Existe una manera fácil de determinar si para cualquier juego dado las estrategias minimax son a prueba de espías. Lo que tenemos que buscar son **puntos de silla**, esto es, pares de estrategias para los cuales el elemento correspondiente en la matriz de resultados es el valor más pequeño en su renglón y el valor más grande en su columna. En el ejemplo 9.5 no hay punto de silla, puesto que el valor más pequeño de cada renglón es también el renglón más pequeño en su columna. Por otra parte, en el juego del ejemplo 9.3 hay un punto de silla que corresponde a las estrategias I y 2 puesto que 8, el valor más pequeño en el segundo renglón, es el valor más grande de la primera columna. También, el juego 3×2 del ejemplo 9.4 tiene un punto de silla que corresponde a las estrategias II y 2 puesto que 3, el valor más pequeño del segundo renglón, es el valor más grande de la segunda columna, y el juego 3×3 en la página 304 tiene un punto de silla que corresponde a las estrategias I y 2 puesto que 2, el valor más pequeño del segundo renglón, es el valor más grande de la primera columna. En general, si un juego tiene un punto de silla, se dice que está **estrictamente determinado** y que las estrategias correspondientes al punto de silla son estrategias minimax a prueba de espías (y por tanto óptimas). En el ejercicio 9.1 se ilustra el hecho que puede haber más de un punto de silla en un juego; también se sigue de este ejercicio que no importa en ese caso cual de los puntos de silla se use para determinar las estrategias óptimas de los dos jugadores.

Si un juego no tiene un punto de silla, las estrategias minimax no son a prueba de espías, y cada jugador puede ser más sagaz que el otro si sabe cómo reaccionará su oponente en una situación dada. Para evitar esta posibilidad, cada jugador debe, de alguna manera, mezclar intencionalmente sus patrones de conducta, y la mejor manera de hacerlo es al introducir un elemento fortuito en la selección de estrategias.

EJEMPLO 9.6

Con respecto al juego del ejemplo 9.5, suponga que el jugador A usa un dispositivo de juego aleatorio (dados, cartas, papeletas numeradas, una tabla de números aleatorios) que lo lleve a escoger la estrategia I con la probabilidad x y a la elección de la estrategia II con probabilidad $1 - x$. Encuentre el valor de x que minimizará la máxima pérdida esperada del jugador A.

Solución

Si la jugadora B escoge la estrategia 1, el jugador A puede esperar perder

$$E = 8x - 5(1 - x)$$

dólares, y si la jugadora B escoge la estrategia 2, el jugador A puede esperar perder

$$E = 2x + 6(1 - x)$$

dólares. Gráficamente, esta situación se describe en la figura 9.1, donde tenemos graficadas las líneas cuyas ecuaciones son $E = 8x - 5(1 - x)$ y $E = 2x + 6(1 - x)$ para los valores de x de 0 a 1.

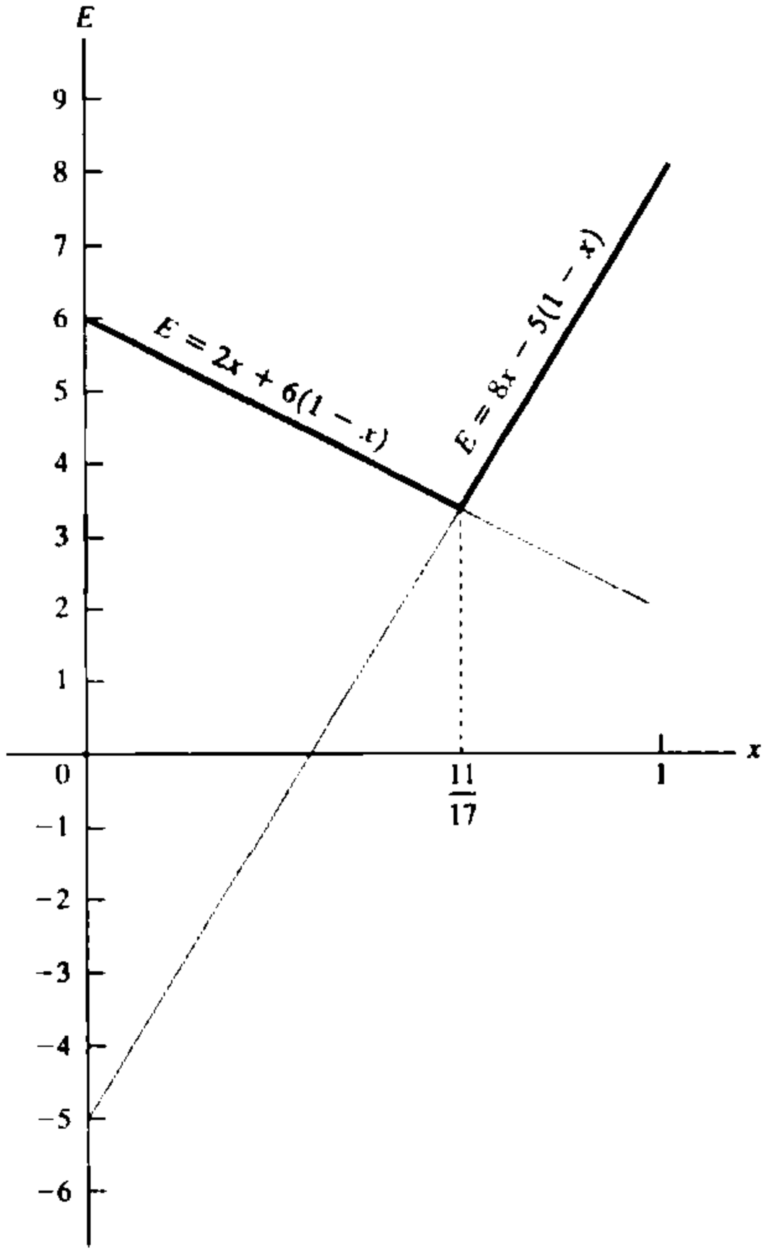

Figura 9.1 Diagrama para el ejemplo 9.6.

Al aplicar el criterio minimax a la pérdida esperada del jugador A, encontramos a partir de la figura 9.1 que el mayor de los dos valores de E para cualquier valor dado de x es el más pequeño donde se intersecan las dos líneas, y para encontrar el valor correspondiente de x, sólo tenemos que resolver la ecuación

$$8x - 5(1 - x) = 2x + 6(1 - x)$$

lo cual da $x = \frac{11}{17}$. Así, si el jugador A usa 11 papeletas numeradas I y seis papeletas numeradas II, las revuelve totalmente, y luego actúa según la clase que saca en forma aleatoria, mantendrá su pérdida esperada máxima en $8 \cdot \frac{11}{17} - 5 \cdot \frac{6}{17} = 3\frac{7}{17}$, o \$3.41 al centavo más cercano. ▲

9.10 Con respecto al ejemplo 9.1, suponga que el fabricante tiene la opción de contratar un pronosticador infalible por \$15,000 para determinar de manera segura si habrá una recesión. Con base en la ventaja original de 2 a 1 de que habrá una recesión, ¿valdría la pena que el fabricante gastara estos \$15,000?

9.11 Cada una de las siguientes matrices de resultados (los pagos que el jugador A hace al jugador B) para un juego de dos personas de suma cero. Elimine todas las estrategias dominadas y determine la estrategia óptima para cada jugador así como el valor del juego:

(a)

3	−2
5	7

(b)

14	11
16	−2

(c)

−5	0	3
−6	−3	−3
−12	−1	1

(d)

7	10	8
8	8	11
7	5	9

9.12 Cada una de las siguientes es la matriz de resultados de un juego de dos personas de suma cero. Encuentre el punto de silla (o puntos de silla) y el valor de cada juego:

(a)

−1	5	−2
0	3	1
−2	−4	5

(b)

3	2	4	9
4	4	4	3
5	6	5	6
5	7	5	9

9.13 Una ciudad pequeña tiene dos gasolineras, las cuales comparten el mercado de gasolina de la ciudad. La dueña de la gasolinera A está considerando si debe regalar vasos a sus clientes como parte de un esquema promocional, y el propietario de la gasolinera B está considerando si debe regalar cuchillos para filete. Ambos saben (por situaciones semejantes en otras partes) que si la estación A regala vasos y la estación B no regala los cuchillos para filete, la participación de mercado de la gasolinera A aumentará en 6 por ciento; si la gasolinera B regala cuchillos para filete y la gasolinera A no regala vasos, la participación de mercado de la gasolinera B aumentará en 8 por ciento; y si ambas estaciones regalan los artículos respectivos, la participación de mercado de la gasolinera B aumentará en 3 por ciento.

(a) Presente esta información en forma de una tabla de resultados donde los elementos son las pérdidas de la gasolinera A en su participación de mercado.

(b) Encuentre las estrategias óptimas para los propietarios de estas dos gasolineras.

9.14 Verifique las dos probabilidades $\frac{4}{17}$ y $\frac{13}{17}$, que dimos en la página 308, para la estrategia aleatoria del jugador B.

que debe recordar al lector del esquema de la página 302. Ahora, θ_1 es el "estado de la Naturaleza" que la moneda tenga dos caras, θ_2 es el "estado de la Naturaleza" que la moneda esté balanceada con cara de un lado y cruz en el otro, a_1 es la decisión del estadístico que la moneda tenga dos caras y a_2 es la decisión del estadístico que la moneda está balanceada con cara de un lado y cruz en el otro. Los elementos en la tabla son los valores correspondientes de la función de pérdida dada.

Ahora consideremos también el hecho que nosotros (jugador A, o el estadístico) sabemos lo que sucedió en el lanzamiento al aire de la moneda; esto es, sabemos si una variable aleatoria X ha asumido el valor de $x = 0$ (cara) o $x = 1$ (cruz). Puesto que queremos hacer uso de esta información al escoger entre a_1 y a_2, necesitamos una función, una **función de decisión**, que nos diga qué acción tomar cuando $x = 0$ y qué acción tomar cuando $x = 1$. Una posibilidad es escoger a_1 cuando $x = 0$ y a_2 cuando $x = 1$, y podemos expresar esto simbólicamente si escribimos

$$d_1(x) = \begin{cases} a_1 & \text{cuando } x = 0 \\ a_2 & \text{cuando } x = 1 \end{cases}$$

o más simplemente, $d_1(0) = a_1$ y $d_1(1) = a_2$. El propósito del subíndice es distinguir esta función de decisión de otras, por ejemplo, de

$$d_2(0) = a_1 \quad \text{y} \quad d_2(1) = a_1$$

que nos dice que escojamos a_1 sin importar el resultado del experimento, de

$$d_3(0) = a_2 \quad \text{y} \quad d_3(1) = a_2$$

que nos dice que escojamos a_2 sin importar el resultado del experimento, y de

$$d_4(0) = a_2 \quad \text{y} \quad d_4(1) = a_1$$

que nos dice que escojamos a_2 cuando $x = 0$ y a_1 cuando $x = 1$.

Para comparar los méritos de todas estas funciones de decisión, determinemos primero las pérdidas esperadas a las cuales nos llevan debido a las diversas estrategias de la Naturaleza, esto es, los valores de la **función de riesgo**

$$R(d_i, \theta_j) = E\{L[d_i(X), \theta_j]\}$$

donde la esperanza se toma con respecto a la variable X. Puesto que las probabilidades para $x = 0$ y $x = 1$ son, respectivamente, 1 y 0 para θ_1 y $\frac{1}{2}$, y $\frac{1}{2}$ para θ_2, obtenemos

$$R(d_1, \theta_1) = 1 \cdot L(a_1, \theta_1) + 0 \cdot L(a_2, \theta_1) = 1 \cdot 0 + 0 \cdot 1 = 0$$

$$R(d_1, \theta_2) = \frac{1}{2} \cdot L(a_1, \theta_2) + \frac{1}{2} \cdot L(a_2, \theta_2) = \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot 0 = \frac{1}{2}$$

$$R(d_2, \theta_1) = 1 \cdot L(a_1, \theta_1) + 0 \cdot L(a_1, \theta_1) = 1 \cdot 0 + 0 \cdot 0 = 0$$

$$R(d_2, \theta_2) = \frac{1}{2} \cdot L(a_1, \theta_2) + \frac{1}{2} \cdot L(a_1, \theta_2) = \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot 1 = 1$$

$$R(d_3, \theta_1) = 1 \cdot L(a_2, \theta_1) + 0 \cdot L(a_2, \theta_1) = 1 \cdot 1 + 0 \cdot 1 = 1$$

$$R(d_3, \theta_2) = \frac{1}{2} \cdot L(a_2, \theta_2) + \frac{1}{2} \cdot L(a_2, \theta_2) = \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 0 = 0$$

$$R(d_4, \theta_1) = 1 \cdot L(a_2, \theta_1) + 0 \cdot L(a_1, \theta_1) = 1 \cdot 1 + 0 \cdot 0 = 1$$

$$R(d_4, \theta_2) = \frac{1}{2} \cdot L(a_2, \theta_2) + \frac{1}{2} \cdot L(a_1, \theta_2) = \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 1 = \frac{1}{2}$$

donde los valores de la función de pérdida se obtuvieron de la tabla en la página 312.

Hemos llegado así al siguiente juego de dos personas de suma cero 4 × 2 en el que los resultados son los valores correspondientes de la función de riesgo:

Jugador A
(*El estadístico*)

Jugador B (*Naturaleza*)		d_1	d_2	d_3	d_4
	θ_1	0	0	1	1
	θ_2	$\frac{1}{2}$	1	0	$\frac{1}{2}$

Como se puede ver mediante una inspección, d_2 está dominada por d_1 y d_4 está dominada por d_3, así que d_2 y d_4 se pueden descartar; en la teoría de decisiones decimos que son **inadmisibles**. En realidad, esto no debe ser una sorpresa puesto que en d_2 así como en d_4 aceptamos la alternativa a_1 (que la moneda tenía dos caras) aun cuando salió cruz.

Esto nos deja con un juego de dos personas de suma cero 2 × 2 en donde el jugador A tiene que escoger entre d_1 y d_3. Se puede verificar fácilmente que si la Naturaleza se considera un oponente malévolo, la estrategia óptima es escoger aleatoriamente entre d_1 y d_3 con probabilidades respectivas de $\frac{2}{3}$ y $\frac{1}{3}$, y el valor del juego (el riesgo esperado) es un $\frac{1}{3}$ de un dólar. Si la Naturaleza no se considera un oponente malévolo, se tendrá que usar algún otro criterio para escoger entre d_1 y d_3, esto se examinará en la sección que sigue. Incidentalmente, formulamos este problema con respecto a una moneda de dos caras y a una moneda normal, pero igualmente podíamos haberlo formulado en forma más abstracta como un problema de decisión en el que tenemos que decidir, con base en una sola observación, si una variable aleatoria tiene la distribución de Bernoulli con el parámetro $\theta = 0$ o el parámetro $\theta = \frac{1}{2}$.

Como ilustración adicional a los conceptos de función de pérdida y función de riesgo, consideremos el siguiente ejemplo, en el cual la Naturaleza al igual que el estadístico tienen un continuo de estrategias.

EJEMPLO 9.7

Una variable aleatoria tiene la densidad uniforme

$$f(x) = \begin{cases} \dfrac{1}{\theta} & \text{para } 0 < x < \theta \\ 0 & \text{en cualquier otra parte} \end{cases}$$

9.5 EL CRITERIO MINIMAX

Si aplicamos el criterio minimax a la ilustración de la sección 9.3, que trata de la moneda ya sea de dos caras o balanceada con cara en un lado y cruz en el otro, encontramos de la tabla en la página 314 con d_2 y d_4 eliminadas que para d_1 el riesgo máximo es $\frac{1}{2}$, y para d_3 el riesgo máximo es 1, y, por tanto, el que minimiza el riesgo máximo es d_1.

EJEMPLO 9.8

Use el criterio minimax para estimar el parámetro θ de una distribución binomial sobre la base de la variable aleatoria X, el número observado de éxitos en n intentos, cuando la función de decisión tiene la forma

$$d(x) = \frac{x + a}{n + b}$$

donde a y b son constantes, y la función de pérdida está dada por

$$L\left(\frac{x + a}{n + b}, \theta\right) = c\left(\frac{x + a}{n + b} - \theta\right)^2$$

donde c es una constante positiva.

Solución

El problema es encontrar los valores de a y b que minimizarán la función de riesgo correspondiente después que se ha maximizado con respecto a θ. Después de todo, tenemos control sobre la elección de a y b, en tanto que la Naturaleza (nuestro supuesto oponente) tiene control sobre la elección de θ.

Puesto que $E(X) = n\theta$ y $E(X^2) = n\theta(1 - \theta + n\theta)$, como vimos en la página 172, se sigue que

$$R(d, \theta) = E\left[c\left(\frac{X + a}{n + b} - \theta\right)^2\right]$$

$$= \frac{c}{(n + b)^2}\left[\theta^2(b^2 - n) + \theta(n - 2ab) + a^2\right]$$

y, al usar el cálculo, podemos encontrar el valor de θ que maximiza esta expresión y entonces minimiza $R(d, \theta)$ para este valor de θ con respecto a a y b. Esto no es especialmente difícil, pero se deja al lector en el ejercicio 9.23 ya que implica ciertos detalles algebraicos tediosos. ▲

Para simplificar el trabajo en un problema de esta clase, a menudo podemos usar el **principio igualador**, según el cual (en condiciones bastante generales) la función de riesgo de la regla de decisión minimax es una constante; por ejemplo, nos dice que en el ejemplo 9.8 la función de riesgo no debe depender del valor de θ.† Para justificar este principio, al menos en forma intuitiva, observe que en el ejemplo 9.6 la estrategia mi-

† Las condiciones exactas en las cuales el principio igualador es válido se encuentran en el libro de T. S. Ferguson que está en las referencias al final de este capítulo

(e) Encuentre la función de decisión que sea la mejor de acuerdo al criterio de Bayes si las probabilidades asignadas a $\theta = \frac{1}{4}$ y $\theta = \frac{1}{2}$ son, respectivamente, $\frac{2}{3}$ y $\frac{1}{3}$.

9.29 Un fabricante produce un artículo que consiste en dos componentes, ambos deben operar para que el artículo funcione correctamente. El costo de regresar uno de los artículos al fabricante para reparaciones es α dólares, el costo de inspeccionar uno de los componentes es β dólares, y el costo de reparar un componente defectuoso es φ dólares. El fabricante puede embarcar cada artículo sin inspeccionar con la garantía de que se dejará en condiciones perfectas de trabajo en su fábrica en caso de que no funcione; puede inspeccionar ambos componentes y repararlos si es necesario; o puede seleccionar aleatoriamente uno de los componentes y embarcar el artículo con la garantía original si trabaja, o repararlo y también verificar el otro componente.

(a) Construya una tabla que muestre la pérdida esperada del fabricante que corresponde a las tres "estrategias" y los tres "estados" de la Naturaleza que 0, 1 o 2 de los componentes no trabajen.

(b) ¿Qué debe hacer el fabricante si $\alpha = \$25.00$, $\varphi = \$10.00$, y quiere minimizar su pérdida máxima esperada?

(c) ¿Qué debe hacer el fabricante para minimizar su riesgo de Bayes si $\alpha = \$10.00$, $\beta = \$12.00$, $\varphi = \$30.00$, y siente que las probabilidades de 0, 1, y 2 componentes defectuosos son, respectivamente, 0.70, 0.20 y 0.10?

REFERENCIAS

Material bastante elemental sobre la teoría de juegos y la teoría de decisiones se puede encontrar en

CHERNOFF, H., and MOSES, L. E., *Elementary Decision Theory*. Mineola, N. Y.: Dover Publications, Inc. (Republication of 1959 edition).

DRESHER, M., *Games of Strategy: Theory and Applications*. Upper Saddle River, N.J.: Prentice Hall, 1961,

HAMBURG, M., *Statistical Analysis for Decision Making*, 4th ed. Orlando, Fla.: Harcourt Brace Jovanovich, 1988,

MCKINSEY, J. C. C., *Introduction to the Theory of Games*. Nueva York: McGraw-Hill Book Company, 1952,

OWEN, G., *Game Theory*. Philadelphia: W. B. Saunders Company, 1968,

WILLIAMS, J. D., *The Compleat Strategyst*. Nueva York: McGraw-Hill Book Company, 1954.

y tratamientos más avanzados en

BICKEL, P. J., and DOKSUM, K. A., *Mathematical Statistics: Basic Ideas and Selected Topics*. Upper Saddle River, N.J.: Prentice Hall, 1977,

FERGUSON, T. S., *Mathematical Statistics: A Decision Theoretic Approach*. Nueva York: Academic Press, Inc., 1967,

WALD, A., *Statistical Decision Functions*. Nueva York: John Wiley & Sons, Inc., 1950.

CAPÍTULO

10

Estimación: teoría

10.1 INTRODUCCIÓN

Tradicionalmente, los problemas de inferencia estadística se dividen en **problemas de estimación** y **pruebas de hipótesis,** aunque en realidad todos son problemas de decisión y, por tanto, se pueden manejar con el enfoque unificado que presentamos en el capítulo anterior. La diferencia principal entre las dos clases de problemas es que en los problemas de estimación debemos determinar el valor de un parámetro (o los valores de varios parámetros) de un continuo posible de alternativas, mientras que en las pruebas de hipótesis debemos decidir si aceptamos o rechazamos un valor específico, o un conjunto de valores específicos, de un parámetro (o los de varios parámetros).

Cuando usamos el valor de una estadística para estimar un parámetro de población, llamamos a esto **estimación puntual**, y nos referimos al valor de la estadística como un **estimador puntual** del parámetro. Por ejemplo, si usamos el valor de $\overline{X}$ para estimar la media de una población, una proporción muestral observada para estimar el parámetro θ de una población binomial, o un valor de S^2 para estimar una varianza de población, en cada caso usamos una estimación puntual del parámetro en cuestión. Estas estimaciones se llaman estimadores puntuales porque en cada caso un número único, o un punto único en el eje real, se usa para estimar el parámetro.

Correspondientemente, nos referimos a las estadísticas mismas como **estimadores puntuales**. Por ejemplo, $\overline{X}$ se puede usar como un estimador puntual de μ, en cuyo caso $\bar{x}$ es un punto estimado de este parámetro. En forma similar, S^2 se puede usar como un estimador puntual de σ^2, en cuyo caso s^2 es un estimador puntual de este parámetro. Aquí usamos la palabra "puntual" para distinguir entre estos estimadores y estimaciones y los **estimadores de intervalos** y las **estimaciones de intervalos**, que presentaremos en el capítulo 11.

Puesto que los estimadores son variables aleatorias, uno de los problemas clave de la estimación puntual es estudiar las distribuciones muestrales. Por ejemplo, cuando estimamos la varianza de una población con base en una muestra aleatoria, difícilmente podemos esperar que el valor de S^2 que obtenemos será realmente igual a σ^2, pero nos tranquilizaría, al menos, saber si podemos esperar que esté cerca. También, debemos decidir si usar una media de la muestra o una mediana de la muestra para estimar la media de una población, sería importante saber, entre otras cosas, si $\overline{X}$ o $\widetilde{X}$ es más probable que nos dé un valor que sea en realidad cercano.

Así, se pueden usar diversas propiedades estadísticas de los estimadores para decidir qué estimador es más apropiado en una situación dada, cuál nos expone a un riesgo más pequeño, cuál nos dará la mayor información al costo más bajo, y así sucesivamente. En particular las propiedades de los estimadores que examinaremos en las secciones 10.2 a 10.6 son **insesgabilidad**, **varianza mínima, eficiencia, consistencia, suficiencia** y **robustez.**

10.2 ESTIMADORES INSESGADOS

Como vimos en la página 315, no existen funciones de decisión perfectas, y en relación con los problemas de estimación esto significa que no hay estimadores perfectos que siempre den la respuesta correcta. Así, parecería razonable que un estimador deba hacerlo al menos en el promedio; esto es, su valor esperado debe ser igual al parámetro que se supone estima. En este caso, se dice que el estimador es **insesgado**; de otra manera, se dice que es **sesgado**. Formalmente:

> **DEFINICIÓN 10.1** Una estadística $\hat{\Theta}$ es un **estimador insesgado** del parámetro $E(\hat{\Theta}) = \theta$.

Los siguientes son algunos ejemplos de estimadores insesgados y sesgados.

EJEMPLO 10.1

Si X tiene la distribución binomial con los parámetros n y θ, demuestre que la proporción muestral, $\frac{X}{n}$, es un estimador insesgado de θ.

Solución

Puesto que $E(X) = n\theta$, se sigue que

$$E\left(\frac{X}{n}\right) = \frac{1}{n}\cdot E(X) = \frac{1}{n}\cdot n\theta = \theta$$

y por tanto que $\frac{X}{n}$ es un estimador insesgado de θ. ▲

EJEMPLO 10.2

Muestre que a menos que $\theta = \frac{1}{2}$, el estimador minimax del parámetro binomial θ en la página 317 es sesgado.

Solución

Puesto que $E(X) = n\theta$, se sigue que

$$E\left(\frac{X + \frac{1}{2}\sqrt{n}}{n + \sqrt{n}}\right) = \frac{E\left(X + \frac{1}{2}\sqrt{n}\right)}{n + \sqrt{n}} = \frac{n\theta + \frac{1}{2}\sqrt{n}}{n + \sqrt{n}}$$

y se puede ver fácilmente que esta cantidad no es igual a θ a menos que $\theta = \frac{1}{2}$. ▲

EJEMPLO 10.3

Si $X_1, X_2, \ldots, X_n$ constituyen una muestra aleatoria de la población dada por

$$f(x) = \begin{cases} e^{-(x-\delta)} & \text{para } x > \delta \\ 0 & \text{en cualquier otra parte} \end{cases}$$

demuestre que $\overline{X}$ es un estimador sesgado de δ.

Solución

Puesto que la media de la población es

$$\mu = \int_{\delta}^{\infty} x \cdot e^{-(x-\delta)}\, dx = 1 + \delta$$

se sigue por el teorema 8.6 que $E(\overline{X}) = 1 + \delta \neq \delta$ y por tanto que $\overline{X}$ es un estimador sesgado de δ. ▲

Cuando $\hat{\Theta}$ es un estimador sesgado de θ, tal vez sea interesante conocer la extensión del **sesgo**, dada por

$$b(\theta) = E(\hat{\Theta}) - \theta$$

Así, para el ejemplo 10.2, el sesgo es

$$\frac{n\theta + \frac{1}{2}\sqrt{n}}{n + \sqrt{n}} - \theta = \frac{\frac{1}{2} - \theta}{\sqrt{n} + 1}$$

y se puede observar que tiende a ser pequeño cuando θ está cercano a $\frac{1}{2}$ y también cuando n es grande. Ciertamente, $\lim_{n\to\infty} b(\theta) = 0$ y decimos que el estimador es **asintóticamente** insesgado.

Por lo que se refiere al ejemplo 10.3, el sesgo es $(1 + \delta) - \delta = 1$, pero en este caso hay algo que podemos hacer al respecto. Puesto que $E(\overline{X}) = 1 + \delta$, se sigue que $E(\overline{X} - 1) = \delta$ y por tanto $\overline{X} - 1$ es un estimador insesgado de δ. Lo que sigue es otro ejemplo donde una pequeña modificación de un estimador nos lleva a un estimador que es insesgado.

$$\operatorname{var}(\hat{\Theta}) \geqq \frac{1}{n \cdot E\left[\left(\frac{\partial \ln f(X)}{\partial \theta}\right)^2\right]}$$

donde $f(x)$ es el valor de la densidad de población en x y n es el tamaño de la muestra aleatoria. Esta desigualdad, la **desigualdad de Cramér-Rao**, nos lleva al siguiente resultado.

TEOREMA 10.2 Si $\hat{\Theta}$ es un estimador insesgado de θ y

$$\operatorname{var}(\hat{\Theta}) = \frac{1}{n \cdot E\left[\left(\frac{\partial \ln f(X)}{\partial \theta}\right)^2\right]}$$

entonces $\hat{\Theta}$ es un estimador insesgado de varianza mínima de θ.

En este caso, la cantidad en el denominador se conoce como la **información** sobre θ que que proporciona la muestra (véase también el ejercicio 10.19). Así, mientras menor sea la varianza, mayor es la información.

EJEMPLO 10.5

Muestre que $\overline{X}$ es un estimador insesgado de varianza mínima de la media μ de una población normal.

Solución

Puesto que

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad \text{para } -\infty < x < \infty$$

se sigue que

$$\ln f(x) = -\ln \sigma\sqrt{2\pi} - \frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2$$

de manera que

$$\frac{\partial \ln f(x)}{\partial \mu} = \frac{1}{\sigma}\left(\frac{x-\mu}{\sigma}\right)$$

y por tanto

$$E\left[\left(\frac{\partial \ln f(X)}{\partial \mu}\right)^2\right] = \frac{1}{\sigma^2} \cdot E\left[\left(\frac{X-\mu}{\sigma}\right)^2\right] = \frac{1}{\sigma^2} \cdot 1 = \frac{1}{\sigma^2}$$

Así,

$$\frac{1}{n \cdot E\left[\left(\frac{\partial \ln f(X)}{\partial \mu}\right)^2\right]} = \frac{1}{n \cdot \frac{1}{\sigma^2}} = \frac{\sigma^2}{n}$$

y puesto que $\overline{X}$ es insesgado y $\text{var}(\overline{X}) = \frac{\sigma^2}{n}$ de acuerdo al teorema 8.1, se sigue que $\overline{X}$ es estimador insesgado de varianza mínima de μ. ▲

Sería erróneo concluir a partir de este ejemplo que $\overline{X}$ es un estimador insesgado de varianza mínima de la media de cualquier población. Ciertamente, en el ejercicio 10.30 se pedirá al lector que verifique que esto no es así para muestras aleatorias de tamaño $n = 3$ de la población uniforme continua con $\alpha = \theta - \frac{1}{2}$ y $\beta = \theta + \frac{1}{2}$.

Como hemos indicado, los estimadores insesgados del mismo parámetro suelen compararse en términos del tamaño de su varianza. Si $\hat{\Theta}_1$ y $\hat{\Theta}_2$ son dos estimadores insesgados del parámetro θ de una población dada y la varianza de $\hat{\Theta}_1$ es menor que la varianza de $\hat{\Theta}_2$, decimos que $\hat{\Theta}_1$ es **relativamente más eficiente** que $\hat{\Theta}_2$. También, usamos la razón

$$\frac{\text{var}(\hat{\Theta}_1)}{\text{var}(\hat{\Theta}_2)}$$

como una medida de la eficiencia de $\hat{\Theta}_2$ relativa a $\hat{\Theta}_1$.

EJEMPLO 10.6

En el ejemplo 10.4 demostramos que si $X_1, X_2, \ldots, X_n$ constituyen una muestra aleatoria de una población uniforme con $\alpha = 0$, entonces $\frac{n+1}{n} \cdot Y_n$ es un estimador insesgado de β.

(a) Demuestre que $2\overline{X}$ también es un estimador insesgado de β.

(b) Compare la eficiencia de estos dos estimadores de β.

Solución

(a) Puesto que la media de la población es $\mu = \frac{\beta}{2}$ de acuerdo al teorema 6.1, se sigue por el teorema 8.1 que $E(\overline{X}) = \frac{\beta}{2}$ y por tanto que $E(2\overline{X}) = \beta$. Así, $2\overline{X}$ es un estimador insesgado de β.

(b) Primero debemos encontrar la varianza de los dos estimadores. Usamos la distribución muestral de Y_n y la expresión para $E(Y_n)$ dada en el ejemplo 10.4, y obtenemos

$$E(Y_n^2) = \frac{n}{\beta^n} \cdot \int_0^\beta y_n^{n+1}\, dy_n = \frac{n}{n+2} \cdot \beta^2$$

y

$$\text{var}(Y_n) = \frac{n}{n+2} \cdot \beta^2 - \left(\frac{n}{n+1} \cdot \beta\right)^2$$

Si dejamos los detalles al lector en el ejercicio 10.27, se puede mostrar que

$$\operatorname{var}\left(\frac{n+1}{n}\cdot Y_n\right) = \frac{\beta^2}{n(n+2)}$$

Puesto que la varianza de la población es $\sigma^2 = \frac{\beta^2}{12}$ de acuerdo al teorema 6.1, se sigue por el teorema 8.1 que $\operatorname{var}(\overline{X}) = \frac{\beta^2}{12n}$ y por tanto que

$$\operatorname{var}(2\overline{X}) = 4\cdot\operatorname{var}(\overline{X}) = \frac{\beta^2}{3n}$$

Por consiguiente, la eficiencia de $2\overline{X}$ relativa a $\frac{n+1}{n}\cdot Y_n$ está dada por

$$\frac{\operatorname{var}\left(\frac{n+1}{n}\cdot Y_n\right)}{\operatorname{var}(2\overline{X})} = \frac{\frac{\beta^2}{n(n+2)}}{\frac{\beta^2}{3n}} = \frac{3}{n+2}$$

y se puede ver que para $n > 1$ el estimador basado en la estadística de *n*ésimo orden, es mucho más eficiente que la otra. Para $n = 10$, por ejemplo, la eficiencia relativa es sólo 25 por ciento, y para $n = 25$ sólo es 11 por ciento. ▲

EJEMPLO 10.7

Cuando la media de una población normal se estima sobre la base de una muestra aleatoria de tamaño $2n + 1$, ¿cuál es la eficiencia de la mediana relativa a la media?

Solución

Por el teorema 8.1 sabemos que $\overline{X}$ es insesgado y que

$$\operatorname{var}(\overline{X}) = \frac{\sigma^2}{2n+1}$$

En lo que se refiere a $\widetilde{X}$ es insesgado en virtud de la simetría de la distribución normal alrededor de su media, y sabemos por el análisis que siguió al teorema 8.17 que para muestras grandes

$$\operatorname{var}(\widetilde{X}) = \frac{\pi\sigma^2}{4n}$$

Así, para muestras grandes, la eficiencia de la mediana relativa a la media es:

refiriéndonos en vez a la población uniforme

$$f(x) = \begin{cases} 1 & \text{para } 0 < x < 1 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

(a) Muestre que $E(X) = \frac{1}{2}$, $E(X^2) = \frac{1}{3}$ y $\text{var}(X) = \frac{1}{12}$ para esta población de manera que para una muestra aleatoria de tamaño $n = 3$, $\text{var}(\overline{X}) = \frac{1}{36}$.

(b) Use los resultados de los ejercicios 8.72 y 8.78 (o derive las densidades y densidad conjunta necesarias) para demostrar que para una muestra aleatoria de tamaño $n = 3$ de esta población, las estadísticas de orden Y_1 y Y_3 tienen $E(Y_1) = \frac{1}{4}$, $E(Y_1^2) = \frac{1}{10}$, $E(Y_3) = \frac{3}{4}$, $E(Y_3^2) = \frac{3}{5}$ y $E(Y_1 Y_3) = \frac{1}{5}$ de manera que $\text{var}(Y_1) = \frac{3}{80}$, $\text{var}(Y_3) = \frac{3}{80}$ y $\text{cov}(Y_1, Y_3) = \frac{1}{80}$.

(c) Use los resultados del inciso (b) y el teorema 4.14 para mostrar que $E\left(\frac{Y_1 + Y_3}{2}\right) = \frac{1}{2}$ y $\text{var}\left(\frac{Y_1 + Y_3}{2}\right) = \frac{1}{40}$ y por tanto que para muestras aleatorias de tamaño $n = 3$ de la población uniforme dada, la mitad de la amplitud es insesgado y más eficiente que la media.

10.31 Demuestre que si $\hat{\Theta}$ es un estimador sesgado de θ, entonces

$$E[(\hat{\Theta} - \theta)^2] = \text{var}(\hat{\Theta}) + [b(\theta)]^2$$

10.32 Si $\hat{\Theta}_1 = \frac{X}{n}$, $\hat{\Theta}_2 = \frac{X+1}{n+2}$ y $\hat{\Theta}_3 = \frac{1}{3}$ son estimadores del parámetro θ de una población binomial y $\theta = \frac{1}{2}$, ¿para qué valores de n

(a) el error medio cuadrático de $\hat{\Theta}_2$ es menor que la varianza de $\hat{\Theta}_1$;

(b) el error medio cuadrático de $\hat{\Theta}_3$ es menor que la varianza de $\hat{\Theta}_1$?

APLICACIONES

10.33 Se toman muestras aleatorias de tamaño n de poblaciones normales con la media μ y las varianzas $\sigma_1^2 = 4$ y $\sigma_2^2 = 9$. Si $\bar{x}_1 = 26.0$ y $\bar{x}_2 = 32.5$, estime μ usando el estimador del inciso (b) del ejercicio 10.21.

10.34 Se toman muestras aleatorias de tamaño n_1 y n_2 de una población normal con la media μ y la varianza σ^2. Si $n_1 = 25$, $n_2 = 50$, $\bar{x}_1 = 27.6$ y $\bar{x}_2 = 38.1$, estime μ usando el estimador del ejercicio 10.23.

10.35 La inteligencia militar de un país sabe que un enemigo construyó ciertos tanques nuevos numerados en serie de 1 a k. Si se capturan tres de estos tanques y sus números de serie son 210, 38 y 155, use el estimador del inciso (b) del ejercicio 10.12 para estimar k.

EJEMPLO 10.10

Si $X_1, X_2, \ldots, X_n$ constituyen una muestra aleatoria de tamaño n de una población de Bernoulli, demuestre que

$$\hat{\Theta} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

es un estimador suficiente del parámetro θ.

Solución

Por la definición 5.2

$$f(x_i;\theta) = \theta^{x_i}(1-\theta)^{1-x_i} \qquad \text{para } x_i = 0, 1$$

de manera que

$$\begin{aligned} f(x_1, x_2, \ldots, x_n) &= \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} \\ &= \theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{n-\sum_{i=1}^{n} x_i} \\ &= \theta^{x}(1-\theta)^{n-x} \\ &= \theta^{n\hat{\theta}}(1-\theta)^{n-n\hat{\theta}} \end{aligned}$$

para $x_i = 0$ o 1 e $i = 1, 2, \ldots, n$. También, puesto que

$$X = X_1 + X_2 + \cdots + X_n$$

es una variable aleatoria binomial con los parámetros θ y n, su distribución está dada por

$$b(x; n, \theta) = \binom{n}{x}\theta^{x}(1-\theta)^{n-x}$$

y la técnica de la transformación de la variable de la sección 7.3 nos da

$$g(\hat{\theta}) = \binom{n}{n\hat{\theta}}\theta^{n\hat{\theta}}(1-\theta)^{n-n\hat{\theta}} \qquad \text{para } \hat{\theta} = 0, \frac{1}{n}, \ldots, 1$$

Ahora, al sustituir en la fórmula para $f(x_1, x_2, \ldots, x_n|\hat{\theta})$ en la página 337, obtenemos

$$\frac{f(x_1, x_2, \ldots, x_n, \hat{\theta})}{g(\hat{\theta})} = \frac{f(x_1, x_2, \ldots, x_n)}{g(\hat{\theta})}$$

$$= \frac{\theta^{n\hat{\theta}}(1-\theta)^{n-n\hat{\theta}}}{\binom{n}{n\hat{\theta}}\theta^{n\hat{\theta}}(1-\theta)^{n-n\hat{\theta}}}$$

$$= \frac{1}{\binom{n}{n\hat{\theta}}}$$

$$= \frac{1}{\binom{n}{x}}$$

$$= \frac{1}{\binom{n}{x_1 + x_2 + \cdots + x_n}}$$

para $x_i = 0$ o 1 e $i = 1, 2, \ldots, n$. Evidentemente, esto no depende de θ y, por consiguiente, hemos mostrado que $\hat{\Theta} = \frac{X}{n}$ es un estimador suficiente de θ. ▲

EJEMPLO 10.11

Muestre que $Y = \frac{1}{6}(X_1 + 2X_2 + 3X_3)$ no es un estimador suficiente del parámetro de Bernoulli θ.

Solución

Puesto que debemos demostrar que

$$f(x_1, x_2, x_3|y) = \frac{f(x_1, x_2, x_3, y)}{g(y)}$$

no es independiente de θ para algunos valores de X_1, X_2 y X_3, consideremos el caso donde $x_1 = 1$, $x_2 = 1$ y $x_3 = 0$. Así, $y = \frac{1}{6}(1 + 2 \cdot 1 + 3 \cdot 0) = \frac{1}{2}$ y

$$f\left(1, 1, 0 \,|\, Y = \frac{1}{2}\right) = \frac{P\left(X_1 = 1, X_2 = 1, X_3 = 0, Y = \frac{1}{2}\right)}{P\left(Y = \frac{1}{2}\right)}$$

$$= \frac{f(1, 1, 0)}{f(1, 1, 0) + f(0, 0, 1)}$$

donde:

$$f(x_1, x_2, x_3) = \theta^{x_1+x_2+x_3}(1-\theta)^{3-(x_1+x_2+x_3)}$$

para $x_i = 0$ o 1 e $i = 1, 2, 3$. Puesto que $f(1, 1, 0) = \theta^2(1-\theta)$ y $f(0, 0, 1) = \theta(1-\theta)^2$, se sigue que

$$f\left(1, 1, 0 | Y = \frac{1}{2}\right) = \frac{\theta^2(1-\theta)}{\theta^2(1-\theta) + \theta(1-\theta)^2} = \theta$$

y se puede ver que esta probabilidad condicional depende de θ. Hemos mostrado así que $Y = \frac{1}{6}(X_1 + 2X_2 + 3X_3)$ no es un estimador suficiente del parámetro θ de una población de Bernoulli. ▲

Porque puede ser muy tedioso comprobar si una estadística es un estimador suficiente de un parámetro basado directamente en la definición 10.3, suele ser más fácil basarlo mejor en el siguiente **teorema de factorización**.

TEOREMA 10.4 La estadística $\hat{\Theta}$ es un estimador suficiente del parámetro θ si y sólo si la distribución o densidad de probabilidad conjunta de la muestra aleatoria se puede factorizar de manera que

$$f(x_1, x_2, \ldots, x_n; \theta) = g(\hat{\theta}, \theta) \cdot h(x_1, x_2, \ldots, x_n)$$

donde $g(\hat{\theta}, \theta)$ sólo depende de $\hat{\theta}$ y θ, y de $h(x_1, x_2, \ldots, x_n)$ no depende de θ.

En los textos más avanzados se puede encontrar una demostración de este teorema; vea, por ejemplo, el libro de Hogg y Craig incluido entre las referencias al final de este capítulo. En este caso, ilustremos el uso del teorema 10.4 por medio del siguiente ejemplo.

EJEMPLO 10.12

Muestre que $\overline{X}$ es un estimador suficiente de la media μ de una población normal con la varianza conocida σ^2.

Solución

Al hacer uso del hecho que

$$f(x_1, x_2, \ldots, x_n; \mu) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \cdot e^{-\frac{1}{2}\cdot\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2}$$

obtenemos:

$$\sum_{i=1}^{n}(x_i - \mu)^2 = \sum_{i=1}^{n}[(x_i - \bar{x}) - (\mu - \bar{x})]^2$$
$$= \sum_{i=1}^{n}(x_i - \bar{x})^2 + \sum_{i=1}^{n}(\bar{x} - \mu)^2$$
$$= \sum_{i=1}^{n}(x_i - \bar{x})^2 + n(\bar{x} - \mu)^2$$

y que

$$f(x_1, x_2, \ldots, x_n; \mu) = \left\{\frac{\sqrt{n}}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right)^2}\right\}$$
$$\times \left\{\frac{1}{\sqrt{n}}\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^{n-1} \cdot e^{-\frac{1}{2}\cdot\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{\sigma}\right)^2}\right\}$$

donde el primer factor del lado derecho depende sólo del estimado $\bar{x}$ y de la media de la población μ, y el segundo factor no implica a μ. De acuerdo al teorema 10.4, se sigue que $\bar{X}$ es un estimador suficiente de la media μ de una población normal con la varianza conocida σ^2. ▲

Con base en la definición 10.3 y el teorema 10.4, respectivamente, hemos presentado dos maneras de comprobar si una estadística $\hat{\Theta}$ es un estimador suficiente de un parámetro dado θ. Como ya dijimos, el teorema de factorización suele llevarnos a soluciones más fáciles; pero si queremos mostrar que una estadística $\hat{\Theta}$ no es un estimador suficiente de un parámetro dado θ, casi siempre es más fácil proceder con la definición 10.3. Esto se ilustró con el ejemplo 10.11.

Mencionemos ahora la siguiente propiedad importante de los estimadores suficientes. Si $\hat{\Theta}$ es un estimador suficiente de θ, entonces cualquier función unívoca $Y = u(\hat{\Theta})$, que no implique a θ, también es un estimador suficiente de θ, y por consiguiente de $u(\theta)$, con tal qu $y = u(\hat{\theta})$ se pueda resolver para dar la inversa unívoca $\hat{\theta} = w(y)$. Esto sigue directamente del teorema 10.4, puesto que podemos escribir

$$f(x_1, x_2, \ldots, x_n; \theta) = g[w(y), \theta] \cdot h(x_1, x_2, \ldots, x_n)$$

donde $g[w(y), \theta]$ depende sólo de y y θ. Si aplicamos este resultado al ejemplo 10.10, donde mostramos que $\hat{\Theta} = \frac{X}{n}$ es un estimador suficiente del parámetro de Bernoulli θ, se sigue que $X = X_1 + X_2 + \cdots + X_n$ también es un estimador suficiente de la media $\mu = n\theta$ de una población binomial.

10.6 ROBUSTEZ

En años recientes, se ha dado atención especial a una propiedad estadística llamada **robustez**. Es indicativa del grado al cual los procedimientos de estimación (y, como veremos más adelante, otros métodos de inferencia) son afectados adversamente por violaciones de las suposiciones que los sustentan. En otras palabras, un estimador se dice que es **robusto** si su distribución muestral no se ve seriamente afectada por violaciones de las suposiciones. Frecuentemente tales violaciones son debidas a puntos extremos causados por errores directos, digamos, al leer los instrumentos o al registrar los datos, o por errores en los procedimientos experimentales. También pueden relacionarse con la naturaleza de las poblaciones muestreadas o con sus parámetros. Por ejemplo, cuando estimamos la vida útil promedio de un cierto componente electrónico, podemos pensar que estamos muestreando una población exponencial, mientras que en realidad estamos muestreando una población Weibull, o cuando estimamos el ingreso promedio de un cierto grupo de edad, podemos usar un método basado sobre la suposición que estamos muestreando una población normal, mientras que en realidad la población (distribución de ingreso) es altamente desviada. También, cuando estimamos la diferencia entre los pesos promedio de dos clases de ranas, la diferencia entre las medias de IQ's de dos grupos étnicos, y en general la diferencia $\mu_1 - \mu_2$ entre las medias de dos poblaciones, podemos suponer que las dos poblaciones tienen la misma varianza σ^2, mientras que en realidad $\sigma_1^2 \neq \sigma_2^2$.

Como debe ser aparente, la mayoría de las preguntas de robustez son difíciles de contestar; ciertamente, mucho del lenguaje usado en el párrafo anterior es relativamente impreciso. Después de todo, ¿qué es lo que queremos decir por "no afectados seriamente"? Es más, cuando hablamos de violaciones de las suposiciones que la sustentan, debe quedar claro que algunas violaciones son mas graves que otras. Cuando se trata de preguntas de robustez, estamos enfrentados así a toda clase de dificultades, matemáticamente y de otro tipo, y por la mayor parte sólo se pueden resolver con simulaciones por computadora. Mencionaremos, otra vez, el tema de la robustez brevemente en la sección 16.1.

EJERCICIOS

10.36 Use la definición 10.2 para mostrar que Y_1, la estadística de primer orden, es un estimador consistente del parámetro α de la población uniforme con $\beta = \alpha + 1$.

10.37 Con respecto al ejercicio 10.36, use el teorema 10.3 para mostrar que $Y_1 - \frac{1}{n+1}$ es un estimador consistente del parámetro α.

10.38 Con respecto a la población uniforme del ejemplo 10.4, use la definición de consistencia para mostrar que Y_n, la estadística de $n$ésimo orden, es un estimador consistente del parámetro β.

10.39 Si $X_1, X_2, \ldots, X_n$ constituyen una muestra aleatoria de tamaño n de una población exponencial, muestre que $\overline{X}$ es un estimador consistente del parámetro θ.

10.40 Con respecto al ejercicio 10.39, ¿es X_n un estimador consistente del parámetro θ?

10.41 Muestre que el estimador del ejercicio 10.21 es consistente.

10.42 Sustituya "asintóticamente insesgado" en vez de "insesgado" en el teorema 10.3, y muestre que $\frac{X+1}{n+2}$ es un estimador consistente del parámetro binomial θ.

10.43 Sustituya "asintóticamente insesgado" en vez de "insesgado" en el teorema 10.3, use este teorema para rehacer el ejercicio 10.38.

10.44 Para mostrar que un estimador puede ser consistente sin ser insesgado o aun insesgado asintóticamente, considere el siguiente procedimiento de estimación: para estimar la media de una población con la varianza finita σ^2, primero tomamos una muestra aleatoria de tamaño n. Después aleatoriamente sacamos una de n papeletas numeradas de 1 a n, y si el número que sacamos es 2, 3,... o n, usamos como nuestro estimador la media de la muestra aleatoria; de lo contrario, usamos el estimado n^2. Demuestre que este procedimiento de estimación es

(a) consistente;

(b) ni insesgado ni asintóticamente insesgado.

10.45 Si $X_1, X_2, \ldots, X_n$ constituyen una muestra aleatoria de tamaño n de una población exponencial, demuestre que $\overline{X}$ es un estimador suficiente del parámetro θ.

10.46 Si X_1 y X_2 son variables aleatorias independientes que tienen distribuciones binomiales con los parámetros θ y n_1 y θ y n_2, muestre que $\dfrac{X_1 + X_2}{n_1 + n_2}$ es un estimador suficiente de θ.

10.47 En respecto al ejercicio 10.46, es $\dfrac{X_1 + 2X_2}{n_1 + 2n_2}$ un estimador suficiente de θ?

10.48 Después de referirse al ejemplo 10.4, ¿es la estadística de $n$ésimo orden, Y_n, un estimador suficiente del parámetro β?

10.49 Si X_1 y X_2 constituyen una muestra aleatoria de tamaño $n = 2$ de una población de Poisson, muestre que la media de la muestra es un estimador suficiente del parámetro λ.

10.50 Si X_1, X_2 y X_3 constituyen una muestra aleatoria de tamaño $n = 3$ de una población de Bernoulli, muestre que $Y = X_1 + 2X_2 + X_3$ no es un estimador suficiente de θ. (*Sugerencia*: considere los valores especiales de X_1, X_2 y X_3.)

10.51 Si $X_1, X_2, \ldots, X_n$ constituyen un muestra aleatoria de tamaño n de una población geométrica, muestre que $Y = X_1 + X_2 + \cdots + X_n$ es un estimador suficiente del parámetro θ.

10.52 Muestre que el estimador del ejercicio 10.5 es un estimador suficiente de la varianza de una población normal con la media conocida μ.

10.7 EL MÉTODO DE MOMENTOS

Como hemos visto en este capítulo, puede haber muchos estimadores diferentes del mismo parámetro de una población. Por consiguiente, parecería deseable tener algún método general, o métodos, que produjeran estimadores con tantas propiedades deseables como fuera posible. En esta sección y en la sección 10.8 presentaremos dos métodos tales, el **método de momentos**, que es históricamente uno de los métodos más antiguos y el **método de máxima verosimilitud**. Adicionalmente, la **estimación Bayesiana** será tratada brevemente en la sección 10.9 y otro método, el **método de los cuadrados mínimos**, será considerado en el capítulo 14.

$$\hat{\alpha} = \frac{(m_1')^2}{m_2' - (m_1')^2} \quad \text{y} \quad \hat{\beta} = \frac{m_2' - (m_1')^2}{m_1'}$$

Puesto que $m_1' = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x}$ y $m_2' = \frac{\sum_{i=1}^{n} x_i^2}{n}$, podemos escribir

$$\hat{\alpha} = \frac{n\bar{x}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \quad \text{y} \quad \hat{\beta} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n\bar{x}}$$

en términos de las observaciones originales. ▲

En estos ejemplos estuvimos interesados en los parámetros de una población específica. Sin embargo, es importante señalar que cuando los parámetros a ser estimados son los momentos de la población, entonces el método de momentos se puede usar sin ningún conocimiento sobre la naturaleza, o la forma funcional, de la población.

10.8 EL MÉTODO DE MÁXIMA VEROSIMILITUD

En dos artículos publicados a principios de este siglo, R. A. Fisher, el prominente estadístico a quien ya hemos mencionado en la página 286, propuso un método general de estimación llamado el **método de máxima verosimilitud**. También mostró las ventajas de este método al demostrar que producía estimadores suficientes siempre que éstos existieran y que los estimadores de máxima verosimilitud son estimadores asintóticamente insesgados de varianza mínima.

Para ayudar a entender el principio en el cual se basa el método de máxima verosimilitud, supongamos que en el correo de la mañana alguien recibe cuatro cartas, pero desafortunadamente una de ellas se extravía antes de que el destinatario tenga la oportunidad de abrirla. Si, entre las tres cartas restantes, dos contienen facturaciones de tarjetas de crédito y la otra no, ¿cuál podría ser un buen estimado de k, el número total de facturaciones de tarjetas de crédito entre las cuatro cartas recibidas? Claramente, k debe ser dos o tres, y si suponemos que cada carta tiene la misma oportunidad de ser extraviada, encontramos que la probabilidad de los datos observados (dos de las tres cartas restantes contienen facturaciones de tarjetas de crédito) es

$$\frac{\binom{2}{2}\binom{2}{1}}{\binom{4}{3}} = \frac{1}{2}$$

para $k = 2$ y

$$\frac{\binom{3}{2}\binom{1}{1}}{\binom{4}{3}} = \frac{3}{4}$$

para $k = 3$. Por consiguiente, si escogemos como nuestro estimado de k el valor que maximiza la probabilidad de obtener los datos observados, obtenemos $k = 3$. Llamamos a éste **el estimado de máxima verosimilitud** y el método por el cual se obtuvo el método de máxima verosimilitud.

Así, la característica esencial del método de máxima verosimilitud es que examinamos los valores de la muestra y entonces escogemos como nuestros estimados de los parámetros desconocidos los valores para los cuales la probabilidad o la densidad de probabilidad de obtener los valores de la muestra es un máximo. En lo que sigue, nos limitaremos al caso de un parámetro; pero, como veremos en el ejemplo 10.18, la idea general también se aplica cuando hay varios parámetros desconocidos. En el caso discreto, si los valores muestrales observados son $x_1, x_2, \ldots, x_n$, la probabilidad de obtenerlos es

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = f(x_1, x_2, \ldots, x_n; \theta)$$

que es justamente el valor de la distribución de probabilidad conjunta de las muestras aleatorias $X_1, X_2, \ldots, X_n$ en $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$. Puesto que los valores muestrales se han observado y, por consiguiente, son números fijos, consideramos $f(x_1, x_2, \ldots, x_n; \theta)$ como un valor de una función de θ, y nos referimos a esta función como la **función de verosimilitud**. Se aplica una definición análoga cuando la muestra aleatoria viene de una población continua, pero en ese caso $f(x_1, x_2, \ldots, x_n; \theta)$ es el valor de la densidad de probabilidad conjunta de las variables aleatorias $X_1, X_2, \ldots, X_n$ en $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$.

DEFINICIÓN **10.5** Si $x_1, x_2, \ldots, x_n$ son los valores de una muestra aleatoria de una población con el parámetro θ, la **función de verosimilitud** de la muestra está dada por

$$L(\theta) = f(x_1, x_2, \ldots, x_n; \theta)$$

para los valores de θ dentro de un dominio dado. En este caso $f(x_1, x_2, \ldots, x_n; \theta)$ es el valor de la distribución de probabilidad conjunta o de la densidad de probabilidad conjunta de las variables aleatorias $X_1, X_2, \ldots, X_n$ en $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$.

Así, el método de máxima verosimilitud consiste en maximizar la función de verosimilitud con respecto a θ, y nos referimos al valor de θ que maximiza la función de verosimilitud como estimador de máxima verosimilitud de θ.

EJEMPLO 10.15

Dado x "éxitos" en n intentos, encuentre el estimador de máxima verosimilitud del parámetro θ de la distribución binomial correspondiente.

Solución

Para encontrar el valor de θ que maximiza

$$L(\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$$

será conveniente hacer uso del hecho que el valor de θ que maximiza $L(\theta)$ también maximiza:

y al igualar la segunda de estas derivadas parciales a cero y despejar para σ^2 después de sustituir $\mu = \bar{x}$, obtenemos

$$\hat{\sigma}^2 = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2 \qquad \blacktriangle$$

Se debe observar que no mostramos que $\hat{\sigma}$ sea un estimador de máxima verosimilitud de σ, sólo que $\hat{\sigma}^2$ es un estimador de máxima verosimilitud de σ^2. Sin embargo, se puede mostrar (véase la referencia al final de este capítulo) que estimadores de máxima verosimilitud tienen la **propiedad de invarianza** que si $\hat{\Theta}$ es un estimador de máxima verosimilitud de θ y la función dada por $g(\theta)$ es continua, entonces $g(\hat{\Theta})$ también es un estimador de máxima verosimilitud de $g(\theta)$. Se sigue que

$$\hat{\sigma} = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2}$$

que difiere de s en que dividimos entre n en vez de $n - 1$, es un estimador de máxima verosimilitud de σ.

En los ejemplos 10.15, 10.16 y 10.18 maximizamos el logaritmo de la función de máxima verosimilitud en vez de la función misma de máxima verosimilitud, pero esto no es de ninguna manera necesario. Justamente dio la casualidad que fue conveniente en cada caso.

EJERCICIOS

10.53 Si $X_1, X_2, \ldots, X_n$ constituyen una muestra aleatoria de una población con la media μ y la varianza σ^2, use el método de momentos para encontrar los estimadores de μ y σ^2.

10.54 Dada una muestra aleatoria de tamaño n de una población exponencial, use el método de momentos para encontrar un estimador del parámetro θ.

10.55 Dada una muestra aleatoria de tamaño n de una población uniforme con $\alpha = 0$, encuentre un estimador para β por el método de momentos.

10.56 Dada una muestra aleatoria de tamaño n de una población De Poisson, use el método de momentos para encontrar un estimador del parámetro λ.

10.57 Dado una muestra aleatoria de tamaño n de una población beta con $\beta = 1$, use el método de momentos para encontrar una fórmula para estimar el parámetro α.

10.58 Si $X_1, X_2, \ldots, X_n$ constituyen una muestra aleatoria de tamaño n de una población dada por:

$$f(x;\theta) = \begin{cases} \dfrac{2(\theta - x)}{\theta^2} & \text{para } 0 < x < \theta \\ 0 & \text{en cualquier otra parte} \end{cases}$$

encuentre un estimador para θ por el método de momentos.

10.59 Si $X_1, X_2, \ldots, X_n$ constituyen una muestra aleatoria de tamaño n de una población dada por:

$$g(x;\theta) = \begin{cases} \frac{1}{\theta} \cdot e^{-\frac{x-\delta}{\theta}} & \text{para } x > \delta \\ 0 & \text{en cualquier otra parte} \end{cases}$$

encuentre los estimadores para δ y θ por el método de momentos. Esta distribución se conoce algunas veces como la **distribución exponencial de dos parámetros**, y para $\theta = 1$ es la distribución del ejemplo 10.3.

10.60 Dada una muestra aleatoria de tamaño n de una población uniforme continua, use el método de momentos para encontrar fórmulas para estimar los parámetros α y β.

10.61 Considere N variables aleatorias independientes que tienen distribuciones binomiales idénticas con los parámetros θ y $n = 3$. Si n_0 de ellas asumen el valor 0, n_1 asumen el valor 1, n_2 asumen el valor 2 y n_3 asumen el valor 3, use el método de momentos para encontrar una fórmula para estimar θ.

10.62 Use el método de máxima verosimilitud para rehacer el ejercicio 10.56.

10.63 Use el método de máxima verosimilitud para rehacer el ejercicio 10.57.

10.64 Si $X_1, X_2, \ldots, X_n$ constituyen una muestra aleatoria de tamaño n de una población gamma con $\alpha = 2$, use el método de máxima verosimilitud para encontrar una fórmula para estimar β.

10.65 Dada una muestra aleatoria de tamaño n de una población normal con la media conocida μ, encuentre el estimador de máxima verosimilitud para σ.

10.66 Si $X_1, X_2, \ldots, X_n$ constituyen una muestra aleatoria de tamaño n para una población geométrica, encuentre las fórmulas para estimar su parámetro θ al usar

(a) el método de momentos;

(b) el método de máxima verosimilitud.

10.67 Dada una muestra aleatoria de tamaño n de una población de Rayleigh (véase el ejercicio 6.20), encuentre un estimador para su parámetro α por el método de máxima verosimilitud.

10.68 Dada una muestra aleatoria de tamaño n de una población Pareto (véase el ejercicio 6.21), use el método de máxima verosimilitud para encontrar una fórmula para estimar su parámetro α.

10.69 Use el método de máxima verosimilitud para rehacer el ejercicio 10.59.

10.70 Use el método de máxima verosimilitud para rehacer el ejercicio 10.60.

10.71 Use el método de máxima verosimilitud para rehacer el ejercicio 10.61.

10.72 Dada una muestra aleatoria de tamaño n de una población gamma con el parámetro conocido α, encuentre un estimador de máxima verosimilitud para

(a) β; (b) $\tau = (2\beta - 1)^2$.

10.73 Si $V_1, V_2, \ldots, V_n$ y $W_1, W_2, \ldots, W_n$ son muestras aleatorias independientes de tamaño n de poblaciones normales con las medias $\mu_1 = \alpha + \beta$ y $\mu_2 = \alpha - \beta$ y la varianza común $\sigma^2 = 1$, encuentre los estimadores de máxima verosimilitud para α y β.

10.74 Si $V_1, V_2, \ldots, V_{n_1}$ y $W_1, W_2, \ldots, W_{n_2}$ son muestras aleatorias independientes de tamaño n_1 y n_2 de poblaciones normales con las medias μ_1 y μ_2, y la varianza común σ^2, encuentre los estimadores de máxima verosimilitud para μ_1, μ_2 y σ^2.

10.75 Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de tamaño n de una población uniforme dada por

$$f(x;\theta) = \begin{cases} 1 & \text{para } \theta - \dfrac{1}{2} < x < \theta + \dfrac{1}{2} \\ 0 & \text{en cualquier otra parte} \end{cases}$$

Muestre que si Y_1 y Y_n son las estadísticas de primer y $n$ésimo orden, cualquier estimador $\hat{\Theta}$ tal que

$$Y_n - \frac{1}{2} \leq \hat{\Theta} \leq Y_1 + \frac{1}{2}$$

puede servir como un estimador de máxima verosimilitud de θ. Esto demuestra que los estimadores de máxima verosimilitud no son únicos.

10.76 Con respecto al ejercicio 10.75, compruebe si los siguientes estimadores son estimadores de máxima verosimilitud de θ:

(a) $\frac{1}{2}(Y_1 + Y_n)$; (b) $\frac{1}{3}(Y_1 + 2Y_2)$.

APLICACIONES

10.77 En 12 días, seleccionados al azar, el consumo de electricidad de una ciudad fue de 6.4, 4.5, 10.8, 7.2, 6.8, 4.9, 3.5, 16.3, 4.8, 7.0, 8.8 y 5.4 millones de kilovatios-hora. Suponiendo que estos datos se pueden considerar como una muestra aleatoria de una población gamma, use los estimadores obtenidos en el ejemplo 10.14 para estimar los parámetros α y β.

10.78 El tamaño de un población animal a veces se estima por el **método de captura-recaptura**. En este método, n_1 de los animales se capturan en el área en consideración, se marcan, y se liberan. Más tarde, se capturan, n_2 de los animales, se encuentra que X de ellos están marcados, y se usa esta información para estimar N, el número total de animales de la clase dada en el área en consideración. Si se capturan $n_1 = 3$ búhos raros en una sección de un bosque, se marcan y se liberan, más tarde se capturan $n_2 = 4$ búhos tales se capturan y sólo uno de ellos se encuentra marcado, estime N por el método de máxima verosimilitud. (*Sugerencia:* intente con $N = 9$, 10, 11, 12, 13 y 14.)

10.79 Ciertos neumáticos radiales tuvieron vidas útiles de 35,200, 41,000, 44,700, 38,600 y 41,500 millas. Suponiendo que estos datos se pueden considerar como una muestra aleatoria de una población exponencial, use el estimador obtenido en el ejercicio 10.54 para estimar el parámetro θ.

10.80 De seis medidas del punto de ebullición de un compuesto de silicio, el tamaño del error fue 0.07, 0.03, 0.14, 0.04, 0.08 y 0.03°C. Suponga que estos datos se pueden considerar como una muestra aleatoria de la población del ejercicio 10.58, use el estimador obtenido ahí por el método de momentos para estimar el parámetro θ.

10.81 Sin contar los que fallaron inmediatamente, ciertos focos tuvieron vidas útiles de 415, 433, 489, 531, 466, 410, 479, 403, 562, 422, 475 y 439 horas. Suponga que estos datos se pueden considerar como una muestra aleatoria de una población exponencial de dos parámetros, use los estimadores obtenidos en el ejercicio 10.59 para estimar los parámetros δ y θ.

10.82 Rehaga el ejercicio 10.81, use los estimadores obtenidos en el ejercicio 10.69 por el método de máxima verosimilitud.

10.83 Los datos reunidos durante varios años muestran que cuando una corredora de bolsa llamó a una muestra aleatoria de ocho de sus clientes, obtuvo una señal de ocupado 6.5, 10.6, 8.1, 4.1, 9.3, 11.5, 7.3 y 5.7 por ciento del tiempo. Suponga que estas cifras se pueden considerar como una muestra aleatoria de una población uniforme continua, use los estimadores obtenidos en el ejercicio 10.60 para estimar los parámetros α y β.

10.84 Rehaga el ejercicio 10.83, use los estimadores obtenidos en el ejercicio 10.70.

10.85 Cada vez que el Sr. Jones va al hipódromo apuesta a tres carreras. En una muestra aleatoria de 20 visitas al hipódromo, el perdió todas sus apuestas 11 veces, ganó una vez siete veces y ganó dos veces en dos ocasiones. Si θ es la probabilidad de que ganará una, cualquiera, de sus apuestas, estímela usando el estimador de máxima verosimilitud obtenido en el ejercicio 10.71.

10.86 En una muestra aleatoria de los maestros en un distrito escolar grande, sus salarios anuales fueron $23,900, $21,500, $26,400, $24,800, $33,600, $24,500, $29,200, $36,200, $22,400, $21,500, $28,300, $26,800, $31,400, $22,700 y $23,100. Suponga que estos datos se pueden considerar como una muestra aleatoria de una población de Pareto, use el estimador obtenido en el ejercicio 10.68 para estimar el parámetro α.

10.87 En 20 días muy fríos, una granjera pudo arrancar su tractor en el primer, tercer, quinto, primer, segundo, primer, tercer, séptimo, segundo, cuarto, cuarto, octavo, primer, tercer, sexto, quinto, segundo, primer, sexto y segundo intento. Suponga que estos datos se pueden considerar como una muestra aleatoria de una población geométrica, estime su parámetro θ por cualquiera de los métodos del ejercicio 10.66.

10.88 Los IQ´s de 10 adolescentes que pertenecen a un grupo étnico son 98, 114, 105, 101, 123, 117, 106, 92, 110 y 108, mientras que los de seis adolescentes que pertenecen a otro grupo étnico son 122, 105, 99, 126, 114 y 108. Suponga que estos datos se pueden considerar como muestras aleatorias independientes de poblaciones normales con las medias μ_1 y μ_2 y la varianza común σ^2, estime estos parámetros por medio de estimadores de máxima verosimilitud obtenidos en el ejercicio 10.74.

10.9 ESTIMACIÓN BAYESIANA†

Hasta ahora hemos supuesto, en este capítulo, que los parámetros que queremos estimar son constantes desconocidas; en la estimación bayesiana los parámetros se consideran como variables aleatorias que tienen **distribuciones previas**, que suelen reflejar la fortaleza de las creencias de uno sobre los valores posibles que pueden asumir. En la sección 9.6 ya encontramos un problema de estimación bayesiana: el parámetro era el de una densidad uniforme y su distribución previa era una distribución gamma.

El problema principal de la distribución bayesiana es el de combinar creencias previas sobre un parámetro con evidencias muestrales directas, y en el ejemplo 9.9 conseguimos esto al determinar $\varphi(\theta|x)$, la densidad condicional de Θ dado $X = x$. En contraste a la distribución previa de Θ, esta distribución condicional (que también refleja la evidencia muestral directa) se llama la **distribución posterior** de Θ. En general, si $h(\theta)$ es el valor de la distribución previa de Θ en θ y queremos combinar la información que expresa con la evidencia muestral directa sobre Θ, por ejemplo, el valor de una estadística $W = u(X_1, X_2, \ldots, X_n)$, determinamos la distribución posterior de Θ por medio de la fórmula

$$\varphi(\theta|w) = \frac{f(\theta, w)}{g(w)} = \frac{h(\theta)\cdot f(w|\theta)}{g(w)}$$

En este caso $f(w|\theta)$ es el valor de la distribución muestral de W dado $\Theta = \theta$ en w, $f(\theta, w)$ es el valor de la distribución conjunta de Θ y W en θ y w, y $g(w)$ es el valor de la distribución marginal de W en w. Advierta que la fórmula anterior para $\varphi(\theta|w)$ es, de hecho, una extensión al caso continuo del teorema de Bayes, teorema 2.13. De ahí el término "estimación bayesiana".

Una vez que se ha obtenido la distribución posterior de un parámetro, se puede usar para hacer estimados como en el ejemplo 9.9, o se puede usar para hacer afirmaciones probabilísticas acerca del parámetro, como se ilustrará en el ejemplo 10.20. Aunque el método que hemos descrito tiene amplias aplicaciones, aquí limitaremos nuestro examen a inferencias sobre el parámetro Θ de una población binomial y la media de una población normal; en el ejercicio 10.92 se tratan las inferencias acerca del parámetro de una población de Poisson.

TEOREMA 10.5 Si X es una variable aleatoria binomial y la distribución previa de Θ es una distribución beta con los parámetros α y β, entonces la distribución posterior de Θ dado $X = x$ es una distribución beta con los parámetros $x + \alpha$ y $n - x + \beta$.

† Algunos de los conceptos y lenguaje usados en esta sección se introdujeron en el capítulo 9, el capítulo opcional sobre teoría de decisiones.

Demostración. Para $\Theta = \theta$ tenemos

$$f(x|\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x} \qquad \text{para } x = 0, 1, 2, \ldots, n$$

$$h(\theta) = \begin{cases} \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cdot\Gamma(\beta)}\cdot\theta^{\alpha-1}(1-\theta)^{\beta-1} & \text{para } 0 < \theta < 1 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

y por tanto

$$f(\theta, x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cdot\Gamma(\beta)}\cdot\theta^{\alpha-1}(1-\theta)^{\beta-1} \times \binom{n}{x}\theta^x(1-\theta)^{n-x}$$

$$= \binom{n}{x}\cdot\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cdot\Gamma(\beta)}\cdot\theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1}$$

para $0 < \theta < 1$ y $x = 0, 1, 2, \ldots, n$, y $f(\theta, x) = 0$ en cualquier otra parte. Para obtener la densidad marginal de X, hagamos uso del hecho que la integral de la densidad beta de 0 a 1 es igual a 1; esto es

$$\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,dx = \frac{\Gamma(\alpha)\cdot\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

Así, obtenemos

$$g(x) = \binom{n}{x}\cdot\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cdot\Gamma(\beta)}\cdot\frac{\Gamma(\alpha+x)\cdot\Gamma(n-x+\beta)}{\Gamma(n+\alpha+\beta)}$$

para $x = 0, 1, \ldots, n$, y por tanto

$$\varphi(\theta|x) = \frac{\Gamma(n+\alpha+\beta)}{\Gamma(\alpha+x)\cdot\Gamma(n-x+\beta)}\cdot\theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1}$$

para $0 < \theta < 1$, y $\varphi(\theta|x) = 0$ en cualquier otra parte. Como se puede ver mediante una inspección, ésta es una densidad beta con los parámetros $x + \alpha$ y $n - x + \beta$. ▼

Para hacer uso de este teorema, refirámonos al resultado que (bajo condiciones muy generales) la media de la distribución posterior minimiza el riesgo de Bayes cuando la función de pérdida es cuadrática, esto es, cuando la función de pérdida está dada por

$$L[d(x), \theta] = c[d(x) - \theta]^2$$

donde c es una constante positiva. Advierta que ésta es la función de pérdida que usamos en el ejemplo 9.9. Puesto que la distribución posterior de Θ es una distribución beta con parámetros $x + \alpha$ y $n - x + \beta$, se sigue por el teorema 6.5 que

$$E(\Theta|x) = \frac{x+\alpha}{\alpha+\beta+n}$$

es un valor de un estimador de θ que minimiza el riesgo de Bayes cuando la función de pérdida es cuadrática y la distribución previa de Θ es de la forma dada.

y

$$\frac{1}{\sigma_1^2} = \frac{10}{(42.5)^2} + \frac{1}{(13.4)^2} = 0.0111$$

de manera que $\sigma_1^2 = 90.0$ y $\sigma_1 = 9.5$. Ahora, la respuesta a nuestra pregunta está dada por el área de la región sombreada de la figura 10.1, esto es, el área bajo la curva normal estándar entre

$$z = \frac{700 - 715}{9.5} = -1.58 \quad \text{y} \quad z = \frac{720 - 715}{9.5} = 0.53$$

Así, la probabilidad de que el valor de M esté entre 700 y 720 es 0.4429 + 0.2019 = 0.6448, o aproximadamente 0.645. ▲

0.2019

0.4429

700 715 720 μ

$z = -1.58$ $z = 0.53$

Figura 10.1 Diagrama para el ejemplo 10.20.

EJERCICIOS

10.89 Con los resultados del ejercicio 6.29, muestre que la media de la distribución posterior de Θ dada en la página 354 se puede escribir como

$$E(\Theta | x) = w \cdot \frac{x}{n} + (1 - w) \cdot \theta_0$$

esto es, como una media ponderada de $\frac{x}{n}$ y θ_0, donde θ_0 y σ_0^2 son la media y la varianza de la distribución beta previa de Θ y

$$w = \frac{n}{n + \frac{\theta_0(1 - \theta_0)}{\sigma_0^2} - 1}$$

10.90 En el ejemplo 10.19 la distribución previa del parámetro Θ de la distribución binomial fue una distribución beta con $\alpha = \beta = 40$. Use el teorema 6.5 para encontrar la media y la varianza de esta distribución previa y describa su forma.

10.91 Muestre que la media de la distribución posterior de M dada en el teorema 10.6 se puede escribir como

$$\mu_1 = w \cdot \bar{x} + (1 - w) \cdot \mu_0$$

esto es, como una media ponderada de $\bar{x}$ y μ_0, donde

$$w = \frac{n}{n + \dfrac{\sigma^2}{\sigma_0^2}}$$

10.92 Si X tiene una distribución de Poisson y la distribución previa del parámetro Λ (mayúscula griega *lambda*) es una distribución gamma con los parámetros α y β, muestre que

(a) la distribución posterior de Λ dada por $X = x$ es una distribución gamma con los parámetros $\alpha + x$ y $\dfrac{\beta}{\beta + 1}$;

(b) la media de la distribución posterior de Λ es

$$\mu_1 = \frac{\beta(\alpha + x)}{\beta + 1}$$

APLICACIONES

10.93 Lo producido en una cierta línea de producción se verifica diariamente mediante la inspección de 100 unidades. Durante un periodo largo de tiempo, el proceso ha mantenido un rendimiento de 80 por ciento, esto es, una proporción defectuosa de 20 por ciento, y la variación de la proporción defectuosa de día a día se mide por una desviación estándar de 0.04. Si en un cierto día la muestra contiene 38 unidades defectuosas, encuentre la media de la distribución posterior de Θ como un estimado de la proporción defectuosa de ese día. Suponga que la distribución previa de Θ es una distribución beta.

10.94 Los registros de una universidad (reunidos durante muchos años) muestran que en promedio 74 por ciento de todos los estudiantes de primer ingreso tiene IQ´s de por lo menos 115. Por supuesto, el porcentaje varía un poco año con año, y esta variación se mide por una desviación estándar de 3 por ciento. Si una verificación muestral de 30 alumnos de primer ingreso que entran a la universidad en 1998 mostró que sólo 18 de ellos tenían IQ´s de por lo menos 115, estime la verdadera proporción de estudiantes con IQ´s de por lo menos 115 en esa generación de primer ingreso, use

(a) sólo la información previa;

(b) sólo la información directa;

(c) el resultado del ejercicio 10.89 para combinar la información previa con la información directa.

10.95 Con respecto al ejemplo 10.20, encuentre $P(712 < \text{M} < 725 | \bar{x} = 692)$.

10.96 Un profesor de historia está preparando un examen final que se administrará a un grupo muy grande de estudiantes. Su creencia acerca de la calificación promedio que deben sacar se expresa subjetivamente por una distribución normal con la media $\mu_0 = 65.2$ y la desviación estándar $\sigma_0 = 1.5$.

(a) ¿Qué probabilidad previa asigna el profesor a la calificación promedio real a que esté en alguna parte del intervalo de 0.63 a 0.68?

(b) ¿Qué probabilidad posterior asignaría a este evento si el examen se probará en una muestra aleatoria de 40 estudiantes cuyas calificaciones tienen una media de 72.9 y una desviación estándar de 7.4? Use $s = 7.4$ como un estimado de σ.

10.97 Una gerente de oficina cree que para una cierta clase de negocio el número diario de llamadas telefónicas que se reciben es una variable aleatoria que tiene una distribución de Poisson, cuyo parámetro tiene una distribución gamma previa con $\alpha = 50$ y $\beta = 2$. Si se le dice que en un negocio así se recibieron 112 llamadas telefónicas en un día dado, ¿cuál será su estimado del promedio diario del número de llamadas recibidas en ese negocio en particular si considera

(a) sólo la información previa;

(b) sólo la información directa;

(c) ambas clases de información y la teoría del ejercicio 10.92?

REFERENCIAS

Se examinan diversas propiedades de estimadores suficientes en

LEHMANN, E. L., *Theory of Point Estimation*. Nueva York: John Wiley & Sons, Inc., 1983.

WILKS, S. S., *Mathematical Statistics*. Nueva York: John Wiley & Sons, Inc., 1962.

y se puede encontrar una demostración del teorema 10.4 en

HOGG, R. V., and CRAIG, A. T., *Introduction to Mathematical Statistics*, 5th ed. Upper Saddle River, N.J.: Prentice Hall, 1995.

Se examinan propiedades importantes de los estimadores de máxima verosimilitud en

KEEPING, E. S., *Introduction to Statistical Inference*. Princeton, N.J.: D. Van Nostrand Co., Inc., 1962,

y se puede encontrar una derivación de la desigualdad de Cramér-Rao, así como de las condiciones más generales bajo la cual se aplica, en

RAO, C. R., *Advanced Statistical Methods in Biometric Research*. Nueva York: John Wiley & Sons, Inc., 1952.

CAPÍTULO

11

Estimación: aplicaciones

11.1 INTRODUCCIÓN

En el capítulo 10 nos centramos en estimación puntual. Aunque ésta es una forma común para expresar las estimaciones, deja espacio para muchas preguntas. Por ejemplo, no nos dice en cuánta información se basa la estimación, ni nos dice nada sobre el tamaño posible del error. Así, tal vez habría que completar un estimador puntual $\hat{\theta}$ de θ con el tamaño de la muestra y el valor de var$(\hat{\Theta})$ o con alguna otra información sobre la distribución muestral de $\hat{\Theta}$. Como veremos, nos permitirá evaluar el tamaño posible del error.

Alternativamente, podríamos usar **estimación de intervalo**. Una estimación de intervalo de θ es un intervalo de la forma $\hat{\theta}_1 < \theta < \hat{\theta}_2$, donde $\hat{\theta}_1$ y son valores de variables aleatorias apropiadas $\hat{\Theta}_1$ y $\hat{\Theta}_2$. Por "apropiada" queremos decir

$$P(\hat{\Theta}_1 < \theta < \hat{\Theta}_2) = 1 - \alpha$$

para alguna probabilidad especificada $1 - \alpha$. Para un valor especificado de $1 - \alpha$, nos referimos a $\hat{\theta}_1 < \theta < \hat{\theta}_2$ como **intervalo de confianza** $(1 - \alpha)100\%$ para θ. También, $1 - \alpha$ se llama **grado de confianza**, y los puntos terminales del intervalo, $\hat{\theta}_1$ y $\hat{\theta}_2$, se llaman **límites de confianza** inferior y superior. Por ejemplo, cuando $\alpha = 0.05$, el grado de confianza es 0.95 y obtenemos un intervalo de confianza del 95%.

Debe entenderse que, como los estimadores puntuales, los estimadores de intervalos de un parámetro dado no son únicos. Esto se ilustra en los ejercicios 11.2 y 11.3 y también en la sección 11.2, donde mostramos que, basado una sola muestra aleatoria, hay varios intervalos de confianza para μ, todos tienen el mismo grado de confianza $1 - \alpha$. Como fue el caso en la estimación puntual, los métodos de estimación de intervalo se juzgan por sus diversas propiedades estadísticas. Por ejemplo, una propiedad deseable es que la longitud de un intervalo de confianza de $(1 - \alpha)\,100\%$ sea tan corta como sea posible; otra propiedad deseable es que la longitud esperada, $E(\hat{\Theta}_2 - \hat{\Theta}_1)$ sea tan pequeña como sea posible.

11.2 LA ESTIMACIÓN DE MEDIAS

Para ilustrar cómo se puede evaluar el tamaño posible de los errores en la estimación puntual, supongamos que la media de una muestra aleatoria se va a usar para estimar la media de una población normal con varianza conocida σ^2. Por el teorema 8.4, la distribución muestral de $\overline{X}$ para muestras aleatorias de tamaño n de una población normal con media μ y varianza σ^2 es una distribución normal con

$$\mu_{\bar{x}} = \mu \quad \text{y} \quad \sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$$

Así, podemos escribir

$$P(|Z| < z_{\alpha/2}) = 1 - \alpha$$

donde

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

y $z_{\alpha/2}$ es tal que la integral de la densidad normal estándar de $z_{\alpha/2}$ a ∞ es igual a $\alpha/2$ (véase también el ejercicio 6.62). Se sigue que

$$P\left(|\overline{X} - \mu| < z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

o, en otras palabras, que

> **TEOREMA 11.1** Si $\overline{X}$, la media de una muestra aleatoria de tamaño n de una población normal con la varianza conocida σ^2, se va a usar como un estimador de la media de la población, la probabilidad es $1 - \alpha$ de que el error será menor que $z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$.

EJEMPLO 11.1

Un equipo de expertos en eficiencia intenta usar la media de una muestra aleatoria de tamaño $n = 150$ para estimar el promedio de la aptitud mecánica de los trabajadores de una línea de ensamble en una industria grande (según la mide cierta prueba estan-

EJEMPLO 11.2

Si una muestra aleatoria de tamaño $n = 20$ de una población normal con la varianza $\sigma^2 = 225$ tiene la media $\bar{x} = 64.3$, construya un intervalo de confianza del 95% para la media de la población μ.

Solución

Sustituimos $n = 20$, $\bar{x} = 64.3$, $\sigma = 15$ y $z_{0.025} = 1.96$ en la fórmula del intervalo de confianza del teorema 11.2, y obtenemos

$$64.3 - 1.96 \cdot \frac{15}{\sqrt{20}} < \mu < 64.3 + 1.96 \cdot \frac{15}{\sqrt{20}}$$

que se reduce a

$$57.7 < \mu < 70.9 \qquad \blacktriangle$$

Como señalamos en la página 360, las fórmulas de intervalos de confianza no son únicas. Esto se puede ver al cambiar la fórmula del intervalo de confianza del teorema 11.2 a

$$\bar{x} - z_{2\alpha/3} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/3} \cdot \frac{\sigma}{\sqrt{n}}$$

o a la fórmula del **intervalo de confianza del** $(1 - \alpha)100\%$ **en un sentido**

$$\mu < \bar{x} + z_\alpha \cdot \frac{\sigma}{\sqrt{n}}$$

Alternativamente, podríamos basar un intervalo de confianza para μ en la mediana de la muestra o, digamos, la mitad de la amplitud.

Hablando estrictamente, los teoremas 11.1 y 11.2 requieren que tratemos con una muestra aleatoria de una población normal con la varianza conocida σ^2. Sin embargo, en virtud del teorema del límite central, también se pueden usar estos resultados para muestras aleatorias de poblaciones no normales siempre que n sea suficientemente grande; esto es, $n \geqq 30$. En este caso, podemos sustituir en vez de σ el valor de la desviación estándar de la muestra.

EJEMPLO 11.3

Un diseñador industrial quiere determinar la cantidad promedio de tiempo que tarda un adulto en ensamblar un juguete "fácil de ensamblar". Use los datos siguientes (en minutos), una muestra aleatoria, para construir un intervalo de confianza del 95% para la media de la población muestreada:

17	13	18	19	17	21	29	22	16	28	21	15
26	23	24	20	8	17	17	21	32	18	25	22
16	10	20	22	19	14	30	22	12	24	28	11

Solución

Al sustituir $n = 36$, $\bar{x} = 19.92$, $z_{0.025} = 1.96$ y $s = 5.73$ con σ la fórmula del intervalo de confianza del teorema 11.2, obtenemos:

llos de la marca B tuvieron un contenido promedio de nicotina de 2.7 miligramos con una desviación estándar de 0.7 miligramos. Suponga que los dos conjuntos de datos son muestras aleatorias independientes de poblaciones normales con varianzas iguales, construya un intervalo de confianza del 95% para la diferencia entre las medias de los contenidos de nicotina de las dos marcas de cigarrillos.

Solución

Primero sustituimos $n_1 = 10$, $n_2 = 8$, $s_1 = 0.5$ y $s_2 = 0.7$ en la fórmula de s_p, y obtenemos

$$s_p = \sqrt{\frac{9(0.25) + 7(0.49)}{16}} = 0.596$$

Entonces, al sustituir este valor junto con $n_1 = 10$, $n_2 = 8$, $\bar{x}_1 = 3.1$, $\bar{x}_2 = 2.7$ y $t_{0.025,16} = 2.120$ (de la tabla IV) en la fórmula del intervalo de confianza del teorema 11.5, encontramos que el intervalo requerido de confianza del 95% es

$$(3.1 - 2.7) - 2.120(0.596)\sqrt{\frac{1}{10} + \frac{1}{8}} < \mu_1 - \mu_2$$
$$< (3.1 - 2.7) + 2.120(0.596)\sqrt{\frac{1}{10} + \frac{1}{8}}$$

que se reduce a

$$-0.20 < \mu_1 - \mu_2 < 1.00$$

Así, los límites de 95% de confianza son -0.20 y 1.00 miligramos; pero observe que puesto que esto incluye $\mu_1 - \mu_2 = 0$, no podemos concluir que hay una diferencia real entre los contenidos promedio de nicotina de las dos marcas de cigarrillos. Encontrará más acerca de esto en el capítulo 13. ▲

EJERCICIOS

11.1 Si x es un valor de una variable aleatoria que tiene una distribución exponencial, encuentre k de manera que el intervalo de 0 a kx es un intervalo de confianza del $(1 - \alpha)100\%$ para el parámetro θ.

11.2 Si x_1 y x_2 son los valores de una muestra aleatoria de tamaño 2 de una población que tiene una densidad uniforme con $\alpha = 0$ y $\beta = \theta$, encuentre k de manera que

$$0 < \theta < k(x_1 + x_2)$$

sea un intervalo de confianza del $(1 - \alpha)100\%$ para θ cuando
(a) $\alpha \leqq \frac{1}{2}$; (b) $\alpha > \frac{1}{2}$.

11.3 Si hacemos uso de los métodos de la sección 8.7, se puede demostrar que para una muestra aleatoria de tamaño $n = 2$ de la población del ejercicio 11.2, la distribución de la amplitud de la muestra está dada por

$$f(R) = \begin{cases} \frac{2}{\theta^2}(\theta - R) & \text{para } 0 < R < \theta \\ 0 & \text{en cualquier otra parte} \end{cases}$$

Use este resultado para encontrar c de manera que

$$R < \theta < cR$$

es un intervalo de confianza del $(1 - \alpha)100\%$ para θ.

11.4 Muestre que el intervalo de confianza del $(1 - \alpha)100\%$

$$\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

es más corto que el intervalo de confianza del $(1 - \alpha)100\%$

$$\bar{x} - z_{2\alpha/3} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/3} \cdot \frac{\sigma}{\sqrt{n}}$$

11.5 Muestre que de todos los intervalos de confianza del $(1 - \alpha)100\%$ de la forma

$$\bar{x} - z_{k\alpha} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{(1-k)\alpha} \cdot \frac{\sigma}{\sqrt{n}}$$

el que tiene $k = 0.5$ es el más corto.

11.6 Muestre que si $\bar{x}$ se usa como una estimación puntual de μ y σ es conocida, la probabilidad es $1 - \alpha$ de que $|\bar{x} - \mu|$, el valor absoluto de nuestro error, no excederá una cantidad especificada e cuando

$$n = \left[z_{\alpha/2} \cdot \frac{\sigma}{e} \right]^2$$

(Si resulta que $n < 30$, esta fórmula no se puede usar, a menos que sea razonable suponer que estamos muestreando a partir de una población normal.)

11.7 Modifique el teorema 11.1 de manera que se pueda usar para evaluar el error máximo cuando σ^2 sea desconocida. (Advierta que este método se puede usar sólo después de haber obtenido los datos.)

11.8 Enuncie un teorema análogo al teorema 11.1, lo que nos permite evaluar el error máximo al usar $\bar{x}_1 - \bar{x}_2$ como una estimación de $\mu_1 - \mu_2$ en las condiciones del teorema 11.4.

11.9 Muestre que S_p^2 es un estimador insesgado de σ^2 y encuentre su varianza en las condiciones del teorema 11.5.

11.10 Verifique el resultado de la página 367, el cual expresa T en términos de $\bar{X}_1$, $\bar{X}_2$ y S_p.

APLICACIONES

11.11 Una funcionaria de distrito intenta usar la media de una muestra aleatoria de 150 alumnos de sexto año de un distrito escolar muy grande para estimar la media de la puntuación que todos los alumnos de sexto año en el distrito obtendrían si tomaran cierta prueba de rendimiento aritmético. Si, basada en la experiencia, la funcionaria sabe que $\sigma = 9.4$ para tales datos, ¿qué se puede afirmar con probabilidad de 0.95 acerca del error máximo?

11.12 Con respecto al ejercicio 11.11, suponga que la funcionaria de distrito toma su muestra y obtiene $\bar{x} = 61.8$. Use toda la información dada para construir un intervalo de confianza del 99% para la media de la puntuación de todos los alumnos de sexto año en el distrito.

11.13 Un investigador médico pretende usar la media de una muestra aleatoria de tamaño $n = 120$ para estimar la media de la presión arterial de mujeres de cincuenta años. Si, con base en su experiencia, sabe que $\sigma = 10.5$ mm de mercurio, ¿qué puede afirmar con probabilidad de 0.99 acerca del error máximo?

11.14 Con respecto al ejercicio 11.13, suponga que el investigador toma su muestra y obtiene $\bar{x} = 141.8$ mm de mercurio. Construya un intervalo de confianza del 98% para la media de la presión arterial de mujeres de cincuenta años.

11.15 Un estudio del crecimiento anual de ciertos cactus mostró que 64 de ellos, seleccionados aleatoriamente en una región desértica, crecieron en promedio 52.80 mm con una desviación estándar de 4.5 mm. Construya un intervalo de confianza del 99% para el verdadero promedio de crecimiento anual de la clase dada de cactus.

11.16 Para estimar el tiempo promedio requerido para ciertas reparaciones, un fabricante de automóviles pidió que se tomara el tiempo a 40 mecánicos, una muestra aleatoria, en la ejecución de esta tarea. Si tardaron un promedio de 24.05 minutos con una desviación estándar de 2.68 minutos, ¿qué puede afirmar el fabricante con 95% de confianza sobre el máximo error si usa $\bar{x} = 24.05$ minutos como una estimación de la media del tiempo real requerido para ejecutar las reparaciones dadas?

11.17 Si una muestra constituye una proporción sensible, esto es, más del 5 por ciento de la población de acuerdo a la regla empírica de la página 275, las fórmulas de los teoremas 11.1 y 11.2 deben modificarse al usar la fórmula de la varianza del teorema 8.6 en vez de la del teorema 8.1. Por ejemplo, el error máximo en el teorema 11.1 se vuelve

$$z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$

Use esta modificación para rehacer el ejercicio 11.11, dado que hay 900 alumnos de sexto año en el distrito escolar.

11.18 Use la modificación sugerida en el ejercicio 11.17 para rehacer el ejercicio 11.12, dado que hay 900 alumnos de sexto año en el distrito escolar.

11.19 Un experto en eficiencia quiere determinar la cantidad promedio de tiempo que tarda la cuadrilla de un foso en cambiar un juego de cuatro neumáticos a un auto de carreras. Use la fórmula para n del ejercicio 11.6 para determinar el tamaño de la muestra que se necesita para que el experto en eficiencia pueda afirmar con 95% de probabilidad que la media de la muestra diferirá de μ, la cantidad a ser estimada, en menos de 2.5 segundos. Se sabe por estudios previos que $\sigma = 12.2$ segundos.

11.20 En un estudio sobre hábitos de ver televisión, se desea estimar el promedio del número de horas a la semana que los adolescentes dedican a verla. Si es razonable suponer que $\sigma = 3.2$ horas, ¿qué tan grande necesita ser la muestra de manera que sea posible afirmar con 95% de confianza que la media de la muestra está errada en menos de 20 minutos? (*Sugerencia*: refiérase al ejercicio 11.6.)

donde $\hat{\Theta} = \frac{X}{n}$. Entonces, si sustituimos $\hat{\theta}$ por θ dentro de los radicales, lo que es una aproximación adicional, obtenemos

TEOREMA 11.6 Si X es una variable aleatoria binomial con los parámetros n y θ, n es grande y $\hat{\theta} = \frac{x}{n}$, entonces

$$\hat{\theta} - z_{\alpha/2} \cdot \sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}} < \theta < \hat{\theta} + z_{\alpha/2} \cdot \sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}$$

es un intervalo de confianza aproximado del $(1-\alpha)100\%$ para θ.

EJEMPLO 11.7

En una muestra aleatoria, 136 personas de 400, a quienes se les aplicó una vacuna contra la influenza, experimentaron cierta incomodidad. Construya un intervalo de confianza del 95% para la verdadera proporción que experimentará alguna incomodidad por la vacuna.

Solución

Sustituimos $n = 400$, $\hat{\theta} = \frac{136}{400} = 0.34$ y $z_{0.025} = 1.96$ en la fórmula del intervalo de confianza del teorema 11.6, y obtenemos

$$0.34 - 1.96\sqrt{\frac{(0.34)(0.66)}{400}} < \theta < 0.34 + 1.96\sqrt{\frac{(0.34)(0.66)}{400}}$$

$$0.294 < \theta < 0.386$$

o, al redondear a dos decimales, $0.29 < \theta < 0.39$. ▲

Usamos las mismas aproximaciones que nos llevaron al teorema 11.6, y también podemos escribir

TEOREMA 11.7 Si $\hat{\theta} = \frac{x}{n}$ se usa como un estimador de θ, podemos afirmar con $(1-\alpha)100\%$ de confianza que el error es menor que

$$z_{\alpha/2} \cdot \sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}$$

EJEMPLO 11.8

Se hace un estudio para determinar la proporción de votantes en una comunidad bastante grande que están a favor de la construcción de una planta nuclear. Si 140 de 400 es una variable aleatoria que tiene aproximadamente la distribución normal estándar. Sustituimos esta expresión por Z en $P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$, y llegamos al siguiente resultado

TEOREMA 11.8 Si X_1 es una variable aleatoria binomial con los parámetros n_1 y θ_1, X_2 es una variable aleatoria binomial con los parámetros n_2 y θ_2, n_1 y n_2 son grandes, y $\hat{\theta}_1 = \frac{x_1}{n_1}$ y $\hat{\theta}_2 = \frac{x_2}{n_2}$, entonces

$$(\hat{\theta}_1 - \hat{\theta}_2) - z_{\alpha/2} \cdot \sqrt{\frac{\hat{\theta}_1(1-\hat{\theta}_1)}{n_1} + \frac{\hat{\theta}_2(1-\hat{\theta}_2)}{n_2}} < \theta_1 - \theta_2$$

$$< (\hat{\theta}_1 - \hat{\theta}_2) + z_{\alpha/2} \cdot \sqrt{\frac{\hat{\theta}_1(1-\hat{\theta}_1)}{n_1} + \frac{\hat{\theta}_2(1-\hat{\theta}_2)}{n_2}}$$

es un intervalo de confianza aproximado de $(1 - \alpha)100\%$ para $\theta_1 - \theta_2$.

EJEMPLO 11.9

Si 132 de 200 votantes hombres y 90 de 159 votantes mujeres están a favor de cierto candidato que hace campaña para gobernador de Illinois, encuentre un intervalo de confianza del 99% para la diferencia entre las proporciones reales de votantes hombres y votantes mujeres que están a favor del candidato.

Solución

Sustituimos $\hat{\theta}_1 = \frac{132}{200} = 0.66$, $\hat{\theta}_2 = \frac{90}{150} = 0.60$ y $z_{0.005} = 2.575$ en la fórmula del intervalo de confianza del teorema 11.8, y obtenemos

$$(0.66 - 0.60) - 2.575\sqrt{\frac{(0.66)(0.34)}{200} + \frac{(0.60)(0.40)}{150}} < \theta_1 - \theta_2$$

$$< (0.66 - 0.60) + 2.575\sqrt{\frac{(0.66)(0.34)}{200} + \frac{(0.60)(0.40)}{150}}$$

la que se reduce a

$$-0.074 < \theta_1 - \theta_2 < 0.194$$

Así, estamos 99% seguros de que el intervalo de -0.074 a 0.194 contiene la diferencia entre las proporciones reales de votantes hombres y mujeres que favorecen al candidato. Observe que esto incluye la posibilidad de una diferencia cero entre las dos proporciones. ▲

EJERCICIOS

11.29 Al despejar

$$-z_{\alpha/2} = \frac{x - n\theta}{\sqrt{n\theta(1-\theta)}} \quad \text{y} \quad \frac{x - n\theta}{\sqrt{n\theta(1-\theta)}} = z_{\alpha/2}$$

para θ, muestre que

$$\frac{x + \frac{1}{2}\cdot z_{\alpha/2}^2 \pm z_{\alpha/2}\sqrt{\frac{x(n-x)}{n} + \frac{1}{4}\cdot z_{\alpha/2}^2}}{n + z_{\alpha/2}^2}$$

son los límites con $(1 - \alpha)100\%$ de confianza para θ.

11.30 Use la fórmula del teorema 11.7 para mostrar que podemos estar al menos $(1 - \alpha)100\%$ seguros de que el error que cometemos es menor que e cuando usamos una proporción muestral $\hat{\theta} = \frac{x}{n}$ con

$$n = \frac{z_{\alpha/2}^2}{4e^2}$$

como una estimación de θ.

11.31 Encuentre una fórmula para n análoga a la del ejercicio 11.30 cuando se sabe que θ debe estar en el intervalo entre θ' y θ''.

11.32 Complete los detalles que llevaron de la estadística Z en la página 374, sustituida en $P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$, a la fórmula del intervalo de confianza del teorema 11.8.

11.33 Encuentre una fórmula para el error máximo análoga a la del teorema 11.7 cuando usamos $\hat{\theta}_1 - \hat{\theta}_2$ como una estimación de $\theta_1 - \theta_2$.

11.34 Use el resultado del ejercicio 11.33 para mostrar que cuando $n_1 = n_2 = n$, podemos estar al menos $(1 - \alpha)100\%$ seguros de que el error que cometemos al usar $\hat{\theta}_1 - \hat{\theta}_2$ como una estimación de $\theta_1 - \theta_2$ es menor que e cuando

$$n = \frac{z_{\alpha/2}^2}{2e^2}$$

APLICACIONES

11.35 Una encuesta muestral en un supermercado mostró que 204 de 300 compradores usan regularmente cupones de descuento. Use la fórmula de muestra grande para el intervalo de confianza del teorema 11.6 para construir un intervalo con 95% de confianza para la verdadera proporción correspondiente.

11.36 Con respecto al ejercicio 11.35, ¿qué podemos decir con 99% de confianza sobre el error máximo si usamos la proporción muestral observada como una estimación de la proporción de todos los compradores en la población muestreada que usan cupones de descuento?

11.37 En una muestra aleatoria de 250 telespectadores en una ciudad grande, 190 habían visto cierto programa polémico. Construya un intervalo de confianza del 99% para la verdadera proporción correspondiente, use

(a) la fórmula de muestras grandes para el intervalo de confianza del teorema 11.6;

(b) los límites de confianza del ejercicio 11.29.

11.38 Con respecto al ejercicio 11.37, ¿qué podemos decir con 95% de confianza acerca del error máximo si usamos la proporción observada de la muestra como una estimación de la verdadera proporción correspondiente?

11.39 Entre 100 peces capturados en cierto lago, 18 no eran comestibles como resultado de la contaminación química del ambiente. Construya un intervalo de confianza del 99% para la verdadera proporción correspondiente.

11.40 En una muestra aleatoria de 120 animadoras, 54 habían sufrido daños, de moderados a severos, en sus voces. Con 90% de confianza, ¿qué podemos decir sobre el error máximo si usamos la proporción muestral $\frac{54}{120} = 0.45$ como una estimación de la verdadera proporción de animadoras que padecen de esta manera?

11.41 En una muestra aleatoria de 300 personas que comen en la cafetería de una tienda departamental, sólo 102 pidieron postre. Si usamos $\frac{102}{300} = 0.34$ como una estimación de la verdadera proporción correspondiente, ¿con qué confianza podemos afirmar que nuestro error es menor que 0.05?

11.42 Una política solicita una encuesta de opinión privada para estimar qué proporción de sus electores están a favor de que ciertas violaciones menores de narcóticos ya no constituyan un delito. Use la fórmula del ejercicio 11.30 para determinar qué tan grande deberá ser la muestra de la encuesta para tener al menos 95% de confianza de que la proporción muestral tiene un error menor que 0.02.

11.43 Use el resultado del ejercicio 11.31 para rehacer el ejercicio 11.42, dado que la encuesta tiene razones para creer que la verdadera proporción no excede 0.30.

11.44 Suponga que queremos estimar qué proporción de todos los automovilistas exceden el límite legal de la velocidad en cierto tramo de la carretera entre Los Ángeles y Bakersfield. Use la fórmula del ejercicio 11.30 para determinar de qué tamaño se necesitará la muestra a fin de estar al menos 99% seguro de que la estimación resultante, la proporción muestral, tiene un error de menos de 0.04.

11.45 Use el resultado del ejercicio 11.31 para rehacer el ejercicio 11.44, dado que tenemos buenas razones para creer que la proporción que estamos tratando de estimar es al menos 0.65.

11.46 En una muestra aleatoria de visitantes a un sitio turístico famoso, 84 de 250 hombres y 156 de 250 mujeres compraron recuerdos. Construya un intervalo de confianza del 95% para la verdadera proporción de hombres y mujeres que compran recuerdos en este sitio turístico.

11.47 Entre 500 solicitudes de licencias de matrimonio escogidas aleatoriamente en un año dado, hubieron 48 en que la mujer era al menos un año mayor que el hombre, y entre 400 solicitudes de licencias de matrimonio escogidas aleatoriamente seis años después, hubieron 68 en los cuales la mujer era al menos un año mayor que el hombre. Construya un intervalo de confianza del 99% para la diferencia entre las verdaderas proporciones correspondientes de solicitudes de licencias de matrimonio en que la mujer es al menos un año mayor que el hombre.

11.48 Con respecto al ejercicio 11.47, ¿qué podemos decir con 98% de confianza acerca del error máximo si usamos la diferencia entre las proporciones muestrales observadas como una estimación de la diferencia entre las verdaderas proporciones correspondientes? (*Sugerencia*: use el resultado del ejercicio 11.33.)

11.49 Suponga que queremos determinar la diferencia entre las proporciones de clientes de una cadena de donas en Carolina del Norte y Vermont que prefieren las donas de la cadena a las de todos sus competidores. Use la fórmula del ejercicio 11.34 para determinar el tamaño de las muestras que se necesitan para estar 95% seguros de que la diferencia entre las dos proporciones muestrales está en error por menos de 0.05.

11.6 LA ESTIMACIÓN DE VARIANZAS

Dada una muestra aleatoria de tamaño n de una población normal, podemos obtener un intervalo de confianza del $(1-\alpha)100\%$ para σ^2 al hacer uso del teorema 8.11, de acuerdo al cual

$$\frac{(n-1)S^2}{\sigma^2}$$

es una variable aleatoria que tiene la distribución ji cuadrada con $n-1$ grados de libertad. Así

$$P\left[\chi^2_{1-\alpha/2,\,n-1} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{\alpha/2,\,n-1}\right] = 1-\alpha$$

$$P\left[\frac{(n-1)S^2}{\chi^2_{\alpha/2,\,n-1}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,\,n-1}}\right] = 1-\alpha$$

donde $\chi^2_{\alpha/2,\,n-1}$ y $\chi^2_{1-\alpha/2,\,n-1}$ son como se definen en la página 282, y obtenemos

TEOREMA 11.9 Si s^2 es el valor de la varianza de una muestra aleatoria de tamaño n de una población normal, entonces

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2,\,n-1}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,\,n-1}}$$

es un intervalo de confianza del $(1-\alpha)100\%$ para σ^2.

Los límites correspondientes con $(1-\alpha)100\%$ de confianza para σ se pueden obtener al sacar la raíz cuadrada de los límites de confianza para σ^2.

EJEMPLO 11.10

En 16 corridas de prueba el consumo de gasolina de un motor experimental tuvo una desviación estándar de 2.2 galones. Construya un intervalo de confianza del 99% para σ^2, que mide la verdadera variabilidad del consumo de gasolina del motor.

Solución

Al suponer que los datos observados se pueden considerar como una muestra aleatoria de una población normal, sustituimos $n = 16$ y $s = 2.2$, junto con $\chi^2_{0.005,15} = 32.801$ y $\chi^2_{0.995,15} = 4.601$, obtenidas de la tabla V, en la fórmula del intervalo de confianza del teorema 11.9, y obtenemos

$$\frac{15(2.2)^2}{32.801} < \sigma^2 < \frac{15(2.2)^2}{4.601}$$

o

$$2.21 < \sigma^2 < 15.78 \qquad \blacktriangle$$

Para obtener el intervalo correspondiente con 99% de confianza para σ, sacamos la raíz cuadrada y obtenemos $1.49 < \sigma < 3.97$.

11.7 LA ESTIMACIÓN DE LA RAZÓN O COCIENTE ENTRE DOS VARIANZAS

Si S_1^2 y S_2^2 son las varianzas de muestras aleatorias independientes de tamaño n_1 y n_2 de poblaciones normales, entonces, de acuerdo al teorema 8.15,

$$F = \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2}$$

es una variable aleatoria que tiene una distribución F con $n_1 - 1$ y $n_2 - 1$ grados de libertad. Así, podemos escribir

$$P\left(f_{1-\alpha/2,\, n_1-1,\, n_2-1} < \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2} < f_{\alpha/2,\, n_1-1,\, n_2-1}\right) = 1-\alpha$$

donde $f_{\alpha/2,\, n_1-1,\, n_2-1}$ y $f_{1-\alpha/2,\, n_1-1,\, n_2-1}$ son como se define en la página 287. Puesto que

$$f_{1-\alpha/2,\, n_1-1,\, n_2-1} = \frac{1}{f_{\alpha/2,\, n_2-1,\, n_1-1}}$$

(véase el ejercicio 8.57), se sigue que

TEOREMA 11.10 Si s_1^2 y s_2^2 son los valores de las varianzas de muestras aleatorias independientes de tamaño n_1 y n_2 de poblaciones normales, entonces

$$\frac{s_1^2}{s_2^2} \cdot \frac{1}{f_{\alpha/2,\, n_1-1,\, n_2-1}} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2}{s_2^2} \cdot f_{\alpha/2,\, n_2-1,\, n_1-1}$$

es un intervalo de confianza del $(1-\alpha)100\%$ para $\frac{\sigma_1^2}{\sigma_2^2}$.

Los límites correspondientes con $(1-\alpha)100\%$ de confianza para $\frac{\sigma_1}{\sigma_2}$ se pueden obtener al sacar la raíz cuadrada de los límites de confianza para $\frac{\sigma_1^2}{\sigma_2^2}$.

EJEMPLO 11.11

Con respecto al ejemplo 11.6, encuentre el intervalo de confianza del 98% para $\frac{\sigma_1^2}{\sigma_2^2}$.

Solución

Sustituimos $n_1 = 10$, $n_2 = 8$, $s_1 = 0.5$, $s_2 = 0.7$ y $f_{0.01,9,7} = 6.72$ y $f_{0.01,7,9} = 5.61$ de la tabla VI, y obtenemos

$$\frac{0.25}{0.49} \cdot \frac{1}{6.72} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{0.25}{0.49} \cdot 5.61$$

o

$$0.076 < \frac{\sigma_1^2}{\sigma_2^2} < 2.862$$

Puesto que el intervalo aquí obtenido incluye la posibilidad de que la razón sea 1, no hay evidencia real contra la suposición de varianzas de población iguales en el ejemplo 11.6. ▲

EJERCICIOS

11.50 Si se puede suponer que el parámetro binomial θ asume un valor cercano a cero, a menudo son útiles los límites de confianza superiores de la forma $\theta < C$. Para una muestra aleatoria de tamaño n, el intervalo en un sentido

$$\theta < \frac{1}{2n} \chi^2_{\alpha,\, 2(x+1)}$$

EJEMPLO 11.12

Para estudiar la durabilidad de una nueva pintura para las líneas blancas centrales, un departamento de obras públicas pintó franjas de prueba de un lado a otro de carreteras con mucho tránsito en ocho lugares diferentes, y los contadores electrónicos mostraron que se deterioraban después de que (al ciento más cercano) 142,600, 167,800, 136,500, 108,300, 126,400, 133,700, 162,000 y 149,400 autos cruzaron por encima de ellas. Construya un intervalo de confianza del 95% para la cantidad promedio de tránsito (cruces de autos) que esta pintura puede soportar antes de deteriorarse.

Solución

La impresión de computadora de la figura 11.1 muestra que el intervalo de confianza deseado es

$$124{,}758 < \mu < 156{,}917$$

cruces de autos. También muestra el tamaño de la muestra, la media de los datos, su desviación estándar y el error estándar estimado de la media, SE MEAN, que está dado por $\dfrac{s}{\sqrt{n}}$. ▲

```
MTB  > SET C1
DATA > 142600  167800    136500    108300    126400    133700    162000    149400
MTB  > TINT 95 C1

           N        MEAN     STDEV     SE MEAN          95.0  PERCENT C.I.
C1         8      140837     19228        6798          [redacted]
```

Figura 11.1 Impresión de computadora para el ejemplo 11.12.

Así como se usaron en este ejemplo, las computadoras nos permiten hacer en forma más eficiente (más rápida, más barata y casi automática) lo que antes se hacía por medio de calculadoras de escritorio, calculadoras manuales o aun a mano. Sin embargo, al tratar con una muestra de tamaño $n = 8$, el ejemplo no puede hacer mucha justicia al poder de las computadoras para manejar conjuntos enormes de datos y ejecutar cálculos que ni siquiera se consideraban posibles hasta años recientes. También, nuestro ejemplo no muestra cómo las computadoras pueden resumir tanto la salida como la entrada y los resultados, así como los datos originales, en varias clases de gráficas y cuadros, lo que permite métodos de análisis que no estaban disponibles en el pasado.

Todo esto es importante, pero no hace justicia al efecto fenomenal que las computadoras han tenido en la estadística. Entre otras cosas, las computadoras se pueden usar para tabular o graficar funciones (digamos, las distribuciones, t, F o χ^2) y así darle al investigador un entendimiento claro de los modelos sustentantes y hacerle posible estudiar los efectos de las violaciones de las suposiciones. También es importante el uso de las computadoras para simular los valores de variables aleatorias (esto es, muestreo de toda clase de poblaciones) cuando no es factible un enfoque matemático formal. Esto provee una herramienta importante cuando estudiamos lo apropiado de los modelos estadísticos.

APLICACIONES

11.60 Se examinaron 20 pilotos en un simulador de vuelo y el tiempo que cada uno tardó en concluir cierta acción correctiva se midió en segundos, con los resultados siguientes:

5.2	5.6	7.6	6.8	4.8	5.7	9.0	6.0	4.9	7.4
6.5	7.9	6.8	4.3	8.5	3.6	6.1	5.8	6.4	4.0

Use un programa de computadora para encontrar un intervalo de confianza del 95% para la media del tiempo que se lleva la acción correctiva.

11.61 Las siguientes son las resistencias a la compresión (dadas a las 10 psi más cercanas) de 30 muestras de concreto

4890	4830	5490	4820	5230	4960	5040	5060	4500	5260
4600	4630	5330	5160	4950	4480	5310	4730	4710	4390
4820	4550	4970	4740	4840	4910	4880	5200	5150	4890

Use un programa de computadora para encontrar un intervalo de confianza del 90% para la desviación estándar de estas resistencias a la compresión.

REFERENCIAS

Se da un método general para obtener intervalos de confianza en

MOOD, A. M., GRAYBILL, F. A., and BOES, D. C., *Introduction to the Theory of Statistics*, 3rd ed. Nueva York: McGraw-Hill Book Company, 1974,

y se pueden encontrar criterios adicionales para juzgar los méritos relativos de los intervalos de confianza en

LEHMANN, E. L., *Testing Statistical Hypotheses*. Nueva York: John Wiley & Sons, Inc., 1959,

y en otros textos sobre estadística matemática. En *Biometrika Tables*, a la que hacemos referencia en la página 298, se dan tablas especiales para construir intervalos de confianza del 95% y 98% para proporciones. Para una prueba de la independencia de las variables aleatorias Z y Y en la página 367, véase

BRUNK, H. D., *An Introduction to Mathematical Statistics*, 3rd ed. Lexington, Mass.: Xerox Publishing Co., 1975.

CAPÍTULO

12

Prueba de hipótesis: teoría

12.1 INTRODUCCIÓN

Problemas como cuando un ingeniero tiene que decidir con base en datos muestrales si el verdadero promedio de vida de cierta clase de neumático es, por lo menos, 22,000 millas, cuando un agrónomo tiene que decidir con base en experimentos si una clase de fertilizante produce un rendimiento más alto de frijol de soya que otro, y cuando un fabricante de productos farmacéuticos tiene que decidir con base en muestras si 90 por ciento de todos los pacientes que reciben un nuevo medicamento se recuperarán de cierta enfermedad, se pueden traducir al lenguaje de las **pruebas estadísticas de hipótesis**. En el primer caso podríamos decir que el ingeniero tiene que probar la hipótesis de que θ, el parámetro de una población exponencial, es por lo menos 22,000; en el segundo caso podríamos decir que el agrónomo tiene que decidir si $\mu_1 > \mu_2$, donde μ_1 y μ_2 son las medias de dos poblaciones normales; y en el tercer caso podríamos decir que el fabricante tiene que decidir si θ, el parámetro de una población binomial, es igual a 0.90. En cada caso se debe suponer, por supuesto, que la distribución escogida describe correctamente las condiciones experimentales; esto es la distribución proporciona el **modelo estadístico** correcto.

Como en los ejemplos anteriores, la mayoría de las pruebas estadísticas de hipótesis tienen que ver con los parámetros de las distribuciones, pero algunas veces también tienen que ver con el tipo, o naturaleza, de las distribuciones mismas. Por ejemplo, en el primero de nuestros tres ejemplos el ingeniero tal vez también tendría que decidir si realmente está tratando con la muestra de una población exponencial o si sus datos son valores de variables aleatorias que tienen, digamos, la distribución de Weibull del ejercicio 6.23.

DEFINICIÓN 12.1 Una **hipótesis estadística** es una afirmación o conjetura acerca de la distribución de una o más variables aleatorias. Si una hipótesis estadística especifica completamente la distribución, se conoce como **hipótesis simple**; si no, se conoce como **hipótesis compuesta**.

Una hipótesis simple debe, por consiguiente, especificar no sólo la forma funcional de la distribución subyacente, sino también los valores de todos los parámetros. Así, en el tercero de los ejemplos anteriores, el que trata de la efectividad del nuevo medicamento, la hipótesis $\theta = 0.90$ es simple, suponiendo, claro está, que especificamos el tamaño de la muestra y que la población es binomial. Sin embargo, en el primero de los ejemplos anteriores la hipótesis es compuesta ya que $\theta \geqq 22{,}000$ no asigna un valor específico al parámetro θ.

Para poder construir un criterio apropiado para probar hipótesis estadísticas, es necesario que también formulemos **hipótesis alternativas**. Para ilustrar esto considere el ejemplo que trata de la vida de los neumáticos, podríamos formular la hipótesis alternativa de que el parámetro θ de la población exponencial es menos de 22,000; en el ejemplo que trata con las dos clases de fertilizantes, podríamos formular la hipótesis alternativa $\mu_1 = \mu_2$; y en el ejemplo que trata del nuevo medicamento, podríamos formular la hipótesis alternativa de que el parámetro θ de la población binomial dada es sólo 0.60, que es la tasa de recuperación de la enfermedad sin el nuevo medicamento.

El concepto de hipótesis simples y compuestas también se aplica a las hipótesis alternativas, y en el primer ejemplo podemos decir ahora que estamos probando la hipótesis compuesta $\theta \geqq 22{,}000$ contra la **alternativa compuesta** $\theta < 22{,}000$, donde θ es el parámetro de una población exponencial. De la misma manera, en el segundo ejemplo estamos probando la hipótesis compuesta $\mu_1 > \mu_2$ contra la alternativa compuesta $\mu_1 = \mu_2$, donde μ_1 y μ_2 son las medias de dos poblaciones normales, y en el tercer ejemplo estamos probando la hipótesis simple $\theta = 0.90$ contra la **alternativa simple** $\theta = 0.60$, donde θ es el parámetro de una población binomial para la cual n está dada.

Frecuentemente, los estadísticos formulan como sus hipótesis exactamente lo contrario de lo que quieren demostrar. Por ejemplo, si queremos demostrar que los estudiantes de una escuela tienen un promedio de IQ más alto que los de otra escuela, podríamos formular la hipótesis de que no hay diferencia: la hipótesis $\mu_1 = \mu_2$. Con esta hipótesis sabemos qué esperar, pero éste no sería el caso si formulamos la hipótesis $\mu_1 > \mu_2$, a menos que especifiquemos la diferencia real entre μ_1 y μ_2.

De igual forma, si queremos demostrar que una clase de mineral tiene un porcentaje más alto de contenido de uranio que otra, podríamos formular la hipótesis de que los dos porcentajes son iguales; y si queremos demostrar que hay una mayor variabilidad en la calidad de un producto de la que hay en la calidad de otro, podríamos formular la hipótesis de que no hay diferencia; esto es, $\sigma_1 = \sigma_2$. En vista de las suposiciones de "no hay diferencia", hipótesis como éstas nos llevan al término **hipótesis nula**, pero hoy en día este término sí es válido para cualquier hipótesis que quisiéramos probar.

De forma simbólica, usaremos H_0 para la hipótesis nula que queremos probar y H_1 o H_A para la hipótesis alternativa. Los problemas con más de dos hipótesis, esto es, problemas que incluyen varias hipótesis alternativas, tienden a ser bastante complicados y no los estudiaremos en este libro.

12.2 PRUEBA DE UNA HIPÓTESIS ESTADÍSTICA

La prueba de una hipótesis estadística es la aplicación de un conjunto explícito de reglas para decidir si aceptamos la hipótesis nula o la rechazamos en favor de la hipótesis alternativa. Suponga, por ejemplo, que un estadístico desea probar la hipótesis nula $\theta = \theta_0$ contra la hipótesis alternativa $\theta = \theta_1$. Para tomar una decisión, generará datos muestrales por medio de un experimento y después calculará el valor de una **estadística de prueba**, que le dirá qué acción tomar para cada resultado posible del espacio muestral. El procedimiento de prueba, por consiguiente, divide los valores posibles de la estadística de prueba en dos subconjuntos: una **región de aceptación** para H_0 y una **región de rechazo** para H_0.

El procedimiento recién descrito puede llevar a dos clases de errores. Por ejemplo, si el verdadero valor del parámetro θ es θ_0 y el estadístico incorrectamente concluye que $\theta = \theta_1$, está cometiendo un error que se conoce como un **error de tipo I**. Por otra parte, si el verdadero valor del parámetro θ es θ_1 y el estadístico concluye en forma incorrecta que $\theta = \theta_0$, está cometiendo una segunda clase de error que se conoce como un **error de tipo II**.

DEFINICIÓN 12.2

1. El rechazo de la hipótesis nula cuando es verdadera se llama un **error de tipo I**; la probabilidad de cometer un error de tipo I se denota con α.
2. La aceptación de la hipótesis nula cuando es falsa se llama un **error de tipo II**; la probabilidad de cometer un error de tipo II se denota con β.

Es costumbre referirse a la región de rechazo para H_0 como la **región crítica** de la prueba y a la probabilidad de obtener un valor de la estadística de prueba dentro de la región crítica cuando H_0 es verdad como el **tamaño** de la región crítica. Así, el tamaño de una región crítica es justamente la probabilidad α de cometer un error de tipo I. Esta probabilidad también se llama el **nivel de significancia** de la prueba (véase el análisis de la página 400).

EJEMPLO 12.1

Con respecto a la tercera ilustración en la página 384, suponga que el fabricante del nuevo medicamento quiere probar la hipótesis nula $\theta = 0.90$ contra la hipótesis alternativa $\theta = 0.60$. Su estadística de prueba es X, el número de éxitos observados (recuperaciones) en 20 intentos, y aceptará la hipótesis nula si $x > 14$; de otra manera la rechazará. Encuentre α y β.

Solución

La región de aceptación para la hipótesis nula es $x = 15, 16, 17, 18, 19$ y 20 y, correspondientemente, la región de rechazo (o región crítica) es $x = 0, 1, 2, ..., 14$. Por consiguiente, de la tabla I:

$$\alpha = P(X \leqq 14; \theta = 0.90) = 0.0114$$

y

$$\beta = P(X > 14; \theta = 0.60) = 0.1255 \quad \blacktriangle$$

Un buen procedimiento de prueba es aquel donde ambas α y β son pequeñas, de ese modo nos da una buena oportunidad de tomar la decisión correcta. La probabilidad de un error de tipo II en el ejemplo 12.1 es más bien alta, pero ésta se puede reducir al cambiar en forma apropiada la región crítica. Por ejemplo, si usamos la región de aceptación $x > 15$ en este ejemplo de manera que la región crítica sea $x \leqq 15$, se puede comprobar con facilidad que esto haría $\alpha = 0.0433$ y $\beta = 0.0509$. Así, aunque se ha reducido la probabilidad de un error de tipo II, se ha vuelto más grande la probabilidad de un error de tipo I. La única forma en que podemos reducir las probabilidades de ambos tipos de errores es aumentar el tamaño de la muestra, pero mientras n se mantenga fija, esta relación inversa entre las probabilidades de errores de tipo I y de tipo II es típica de los procedimientos de decisión estadísticos. En otras palabras, si la probabilidad de un tipo de error se reduce, la del otro tipo de error aumenta.

EJEMPLO 12.2

Suponga que queremos probar la hipótesis nula de que la media de una población normal con $\sigma^2 = 1$ es μ_0 contra la hipótesis alternativa de que es μ_1, donde $\mu_1 > \mu_0$. Encuentre el valor de K tal que $\bar{x} > K$ provea una región crítica de tamaño $\alpha = 0.05$ para una muestra aleatoria de tamaño n.

Solución

Al referirnos a la figura 12.1 y la tabla III, encontramos que $z = 1.645$ corresponde al elemento 0.4500 y por tanto que

$$1.645 = \frac{K - \mu_0}{1/\sqrt{n}}$$

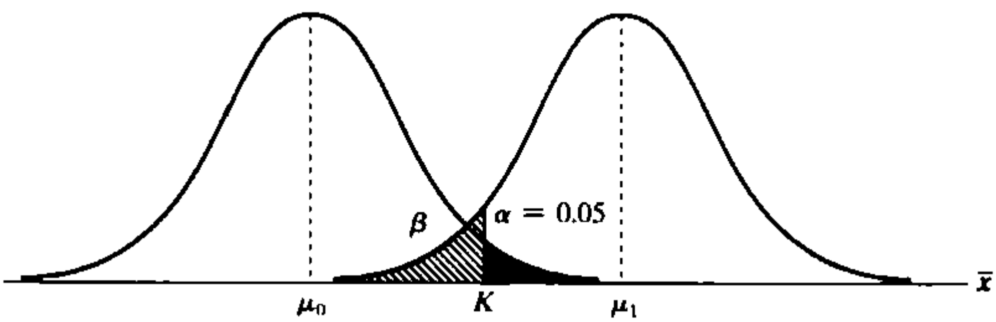

Figura 12.1 Diagrama para los ejemplos 12.2 y 12.3.

Se sigue que

$$K = \mu_0 + \frac{1.645}{\sqrt{n}} \quad \blacktriangle$$

EJEMPLO 12.3

Con respecto al ejemplo 12.2, determine el tamaño mínimo de la muestra necesaria para probar la hipótesis nula $\mu_0 = 10$ contra la hipótesis alternativa $\mu_1 = 11$ con $\beta \leqq 0.06$.

Solución

Puesto que β está dada por el área sombreada en la figura 12.1, obtenemos

$$\begin{aligned}\beta &= P\left(\overline{X} < 10 + \frac{1.645}{\sqrt{n}}; \mu = 11\right)\\ &= P\left[Z < \frac{\left(10 + \frac{1.645}{\sqrt{n}}\right) - 11}{1/\sqrt{n}}\right]\\ &= P(Z < -\sqrt{n} + 1.645)\end{aligned}$$

y puesto que $z = 1.555$ corresponde a un elemento de $0.5000 - 0.06 = 0.4400$ en la tabla III, hacemos $-\sqrt{n} + 1.645$ igual a -1.555. Se sigue que $\sqrt{n} = 1.645 + 1.555 = 3.200$ y $n = 810.24$, u 11 redondeado al entero más cercano. ▲

12.3 PÉRDIDAS Y RIESGOS†

Los conceptos de funciones de pérdida y de riesgo que se introdujeron en el capítulo 9 también juegan una parte importante en la teoría de la prueba de hipótesis. Desde el enfoque de la teoría de decisiones a la prueba de la hipótesis nula de que un parámetro poblacional θ es igual a θ_0 contra la alternativa de que es igual a θ_1, el estadístico bien toma la acción α_0 y acepta la hipótesis nula, o bien toma la acción a_1 y acepta la hipótesis alternativa. Dependiendo del verdadero "estado de la Naturaleza" y de la acción que éste tome, sus pérdidas se muestran en la siguiente tabla:

		Estadístico a_0	*Estadístico* a_1
Naturaleza	θ_0	$L(a_0, \theta_0)$	$L(a_1, \theta_0)$
	θ_1	$L(a_0, \theta_1)$	$L(a_1, \theta_1)$

Estas pérdidas pueden ser positivas o negativas (que reflejan castigos o recompensas), y la única condición que impondremos es que:

† Omita esta sección si se omitió el capítulo 9.

Al probar la hipótesis nula $\theta = \theta_0$ contra la hipótesis alternativa $\theta = \theta_1$, la cantidad $1 - \beta$ se conoce como la **potencia** de la prueba en $\theta = \theta_1$.

Una región crítica para probar una hipótesis nula simple $\theta = \theta_0$ contra una hipótesis alternativa simple $\theta = \theta_1$ se dice que es **mejor** o **más potente**, si la potencia de la prueba en $\theta = \theta_1$ está en un máximo. Para construir una región crítica más potente en esta clase de situación, hacemos referencia a las verosimilitudes (véase la página 346) de una muestra aleatoria de tamaño n de la población en consideración cuando $\theta = \theta_0$ y $\theta = \theta_1$. Denotemos estas verosimilitudes con L_0 y L_1, tenemos así

$$L_0 = \prod_{i=1}^{n} f(x_i; \theta_0) \quad \text{y} \quad L_1 = \prod_{i=1}^{n} f(x_i; \theta_1)$$

Hablando intuitivamente, es evidente que $\frac{L_0}{L_1}$ debe ser pequeña para puntos de la muestra dentro de la región crítica, lo que lleva a errores de tipo I cuando $\theta = \theta_0$ y a decisiones correctas cuando $\theta = \theta_1$; de la misma manera, es evidente que $\frac{L_0}{L_1}$ debe ser grande para puntos de la muestra fuera de la región crítica, lo que lleva a decisiones correctas cuando $\theta = \theta_0$ y a errores de tipo II cuando $\theta = \theta_1$. El siguiente teorema demuestra el hecho que este argumento, ciertamente, garantice una región crítica más potente.

TEOREMA 12.1 *(Lema de Neyman-Pearson)* Si C es una región crítica de tamaño α y k es una constante tal que

$$\frac{L_0}{L_1} \leq k \qquad \text{dentro de } C$$

y

$$\frac{L_0}{L_1} \geq k \qquad \text{fuera de } C$$

entonces C es una región crítica más potente de tamaño α para probar $\theta = \theta_0$ contra $\theta = \theta_1$.

Demostración. Suponga que C es una región crítica que satisface las condiciones del teorema y que D es alguna otra región crítica de tamaño α. Así

$$\int \cdots \int_C L_0 \, dx = \int \cdots \int_D L_0 \, dx = \alpha$$

donde dx representa a $dx_1, dx_2, \ldots, dx_n$, y las dos integrales múltiples se toman sobre las respectivas regiones C y D de n dimensiones. Ahora, al hacer uso del hecho que C es la unión de los conjuntos ajenos $C \cap D$ y $C \cap D'$, mientras que D es la unión de los conjuntos ajenos $C \cap D$ y $C' \cap D$, podemos escribir

$$\int \cdots \int_{C\cap D} L_0 \, dx + \int \cdots \int_{C\cap D'} L_0 \, dx = \int \cdots \int_{C\cap D} L_0 \, dx + \int \cdots \int_{C'\cap D} L_0 \, dx = \alpha$$

y por tanto

$$\int_{C\cap D'} \cdots \int L_0\, dx = \int_{C'\cap D} \cdots \int L_0\, dx$$

Entonces, puesto que $L_1 \geqq L_0/k$ dentro de C y $L_1 \leqq L_0/k$ fuera de C, se sigue que

$$\int_{C\cap D'} \cdots \int L_1\, dx \geqq \int_{C\cap D'} \cdots \int \frac{L_0}{k}\, dx = \int_{C'\cap D} \cdots \int \frac{L_0}{k}\, dx \geqq \int_{C'\cap D} \cdots \int L_1\, dx$$

y por tanto que

$$\int_{C\cap D'} \cdots \int L_1\, dx \geqq \int_{C'\cap D} \cdots \int L_1\, dx$$

Finalmente,

$$\int_{C} \cdots \int L_1\, dx = \int_{C\cap D} \cdots \int L_1\, dx + \int_{C\cap D'} \cdots \int L_1\, dx$$

$$\geqq \int_{C\cap D} \cdots \int L_1\, dx + \int_{C'\cap D} \cdots \int L_1\, dx = \int_{D} \cdots \int L_1\, dx$$

de manera que

$$\int_{C} \cdots \int L_1\, dx \geqq \int_{D} \cdots \int L_1\, dx$$

y esto completa la demostración del teorema 12.1. La desigualdad final enuncia que para la región crítica C la probabilidad de *no* cometer un error de tipo II es mayor que, o igual a, la probabilidad correspondiente para cualquier otra región crítica de tamaño α. (Para el caso discreto la demostración es la misma, donde las sumas toman el lugar de las integrales.) ▼

EJEMPLO 12.4

Una muestra aleatoria de tamaño n de una población normal con $\sigma^2 = 1$ se va a usar para probar la hipótesis nula $\mu = \mu_0$ contra la hipótesis alternativa $\mu = \mu_1$, donde $\mu_1 > \mu_0$. Use el lema de Neyman-Pearson para encontrar la región crítica más potente de tamaño α.

Solución

Las dos verosimilitudes son

$$L_0 = \left(\frac{1}{\sqrt{2\pi}}\right)^n \cdot e^{-\frac{1}{2}\Sigma(x_i-\mu_0)^2} \quad \text{y} \quad L_1 = \left(\frac{1}{\sqrt{2\pi}}\right)^n \cdot e^{-\frac{1}{2}\Sigma(x_i-\mu_1)^2}$$

donde las sumas se extienden de $i = 1$ y $i = n$, y después de algunas simplificaciones su razón se vuelve

$$\frac{L_0}{L_1} = e^{\frac{n}{2}(\mu_1^2 - \mu_0^2) + (\mu_0 - \mu_1)\cdot \Sigma x_i}$$

Así, debemos encontrar una constante k y una región C del espacio muestral tal que

$$e^{\frac{n}{2}(\mu_1^2 - \mu_0^2) + (\mu_0 - \mu_1)\cdot \Sigma x_i} \leqq k \qquad \text{dentro de } C$$

$$e^{\frac{n}{2}(\mu_1^2 - \mu_0^2) + (\mu_0 - \mu_1)\cdot \Sigma x_i} \geqq k \qquad \text{fuera de } C$$

y después de sacar logaritmos, restar $\frac{n}{2}(\mu_1^2 - \mu_0^2)$, y dividir por la cantidad negativa $n(\mu_0 - \mu_1)$, estas dos desigualdades se vuelven

$$\bar{x} \geqq k \qquad \text{dentro de } C$$

$$\bar{x} \leqq k \qquad \text{fuera de } C$$

donde K es una expresión en k, n, μ_0 y μ_1.

En la práctica real, las constantes como K se determinan al hacer uso del tamaño de la región crítica y de la teoría estadística apropiada. En nuestro caso (véase el ejemplo 12.2) obtenemos $K = \mu_0 + z_\alpha \cdot \frac{1}{\sqrt{n}}$, donde z_α es como se define en la página 227. Así, la región crítica más potente de tamaño α para probar la hipótesis nula $\mu = \mu_0$ contra la alternativa $\mu = \mu_1$ (con $\mu_1 > \mu_0$) para la población normal dada es

$$\bar{x} \geqq \mu_0 + z_\alpha \cdot \frac{1}{\sqrt{n}}$$

y se debe observar que no depende de μ_1. Esto es una propiedad importante, a la cual nos volveremos a referir en la sección 12.5. ▲

Advierta que en este caso derivamos la región crítica sin mencionar primero que la estadística de prueba va a ser $\overline{X}$. Puesto que la especificación de una región crítica define así la estadística de prueba correspondiente, y viceversa, estos dos términos "región crítica" y "estadística de prueba", a menudo se usan indistintamente en el lenguaje de la estadística.

EJERCICIOS

12.1 Decida en cada caso si la hipótesis es simple o compuesta:

(a) la hipótesis de que una variable aleatoria tiene una distribución gamma con $\alpha = 3$ y $\beta = 2$;

(b) la hipótesis de que una variable aleatoria tiene una distribución gamma con $\alpha = 3$ y $\beta \neq 2$;

(c) la hipótesis de que una variable aleatoria tiene una densidad exponencial;

(d) la hipótesis de que una variable aleatoria tiene una distribución beta con la media $\mu = 0.50$.

12.2 Decida en cada caso si la hipótesis es simple o compuesta:

(a) la hipótesis de que una variable aleatoria tiene una distribución de Poisson con $\lambda = 1.25$;

(b) la hipótesis de que una variable aleatoria tiene una distribución de Poisson con $\lambda > 1.25$;

(c) la hipótesis de que una variable aleatoria tiene una distribución normal con la media $\mu = 100$;

(d) la hipótesis de que una variable aleatoria tiene una distribución binomial negativa con $k = 3$ y $\theta < 0.60$.

12.3 Una sola observación de una variable aleatoria que tiene una distribución hipergeométrica con $N = 7$ y $n = 2$ se usa para probar la hipótesis nula $k = 2$ contra la hipótesis alternativa $k = 4$. Si la hipótesis nula se rechaza si y sólo si el valor de la variable aleatoria es 2, encuentre las probabilidades de errores tipo I y de tipo II.

12.4 Con respecto al ejemplo 12.1, ¿cuáles hubieran sido las probabilidades de errores de tipo I y de tipo II si la región de aceptación hubiera sido $x > 16$ y la región de rechazo correspondiente hubiera sido $x \leqq 16$?

12.5 Una observación única de una variable aleatoria que tiene una distribución geométrica se usa para probar la hipótesis nula $\theta = \theta_0$ contra la hipótesis alternativa $\theta = \theta_1 > \theta_0$. Si la hipótesis nula se rechaza si y sólo si el valor de la variable aleatoria es mayor que, o igual a, el entero positivo k, encuentre las expresiones para las probabilidades de errores de tipo I y de tipo II.

12.6 Una observación única de una variable aleatoria que tiene una distribución exponencial se usa para probar la hipótesis nula que la media de la distribución es $\theta = 2$ contra la hipótesis alternativa que es $\theta = 5$. Si la hipótesis nula se rechaza si y sólo si el valor de la variable aleatoria es menor que 3, encuentre las probabilidades de errores de tipo I y de tipo II.

12.7 Sea que X_1 y X_2 constituyan una muestra aleatoria de una población normal con $\sigma^2 = 1$. Si la hipótesis nula $\mu = \mu_0$ va a ser rechazada en favor de la hipótesis alternativa $\mu = \mu_1 > \mu_0$ cuando $\bar{x} > \mu_0 + 1$, ¿cuál es el tamaño de la región crítica?

12.8 Una observación única de una variable aleatoria que tiene una densidad uniforme con $\alpha = 0$ se usa para probar la hipótesis nula $\beta = \beta_0$ contra la hipótesis alternativa $\beta = \beta_0 + 2$. Si la hipótesis nula se rechaza si y sólo si el valor de la variable aleatoria asume un valor mayor que $\beta_0 + 1$, encuentre las probabilidades de errores de tipo I y de tipo II.

12.9 Sea que X_1 y X_2 constituyan una muestra aleatoria de tamaño 2 de la población dada por

$$f(x;\theta) = \begin{cases} \theta x^{\theta-1} & \text{para } 0 < x < 1 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

Si la región crítica $x_1 x_2 \geqq \frac{3}{4}$ se usa para probar la hipótesis nula $\theta = 1$ contra la hipótesis alternativa $\theta = 2$, ¿cuál es la potencia de esta prueba en $\theta = 2$?

12.10 Demuestre que si $\mu_1 < \mu_0$ en el ejemplo 12.4, el lema de Neyman-Pearson nos da la región crítica

$$\bar{x} \leqq \mu_0 - z_\alpha \cdot \frac{1}{\sqrt{n}}$$

12.11 Una muestra aleatoria de tamaño n de una población exponencial se usa para probar la hipótesis nula $\theta = \theta_0$ contra la hipótesis alternativa $\theta = \theta_1 > \theta_0$. Use el lema de Neyman-Pearson para encontrar la región crítica más potente de tamaño α, y use el resultado del ejemplo 7.16 para indicar cómo evaluar la constante.

12.12 Use el lema de Neyman-Pearson para indicar cómo construir la región crítica más potente de tamaño α para probar la hipótesis nula $\theta = \theta_0$, donde θ es el parámetro de la distribución binomial con un valor dado de n, contra la hipótesis alternativa $\theta = \theta_1 < \theta_0$.

12.13 Con respecto al ejercicio 12.12, si $n = 100$, $\theta_0 = 0.40$, $\theta_1 = 0.30$ y α es tan grande como sea posible sin exceder de 0.05, use la aproximación normal a la distribución binomial para encontrar la probabilidad de cometer un error de tipo II.

12.14 Una observación única de una muestra aleatoria que tiene una distribución geométrica se va a usar para probar la hipótesis nula que su parámetro es igual a θ_0 contra la hipótesis alternativa que es igual a $\theta_1 > \theta_0$. Use el lema de Neyman-Pearson para encontrar la mejor región crítica de tamaño α.

12.15 Dada una muestra aleatoria de tamaño n de una población normal con $\mu = 0$ use el lema de Neyman-Pearson para construir la región crítica más potente de tamaño α para probar la hipótesis nula $\sigma = \sigma_0$ contra la alternativa $\sigma = \sigma_1 > \sigma_0$.

12.16 Suponga que en el ejemplo 12.1 el fabricante del nuevo medicamento cree que la ventaja es 4 a 1 que con su medicamento la tasa de recuperación de la enfermedad es 0.90 en vez de 0.60. Con esta ventaja, ¿cuáles son las probabilidades que tomará la decisión equivocada si usa la función de decisión

(a) $d_1(x) = \begin{cases} a_0 & \text{para } x > 14 \\ a_1 & \text{para } x \leqq 14 \end{cases}$

(b) $d_2(x) = \begin{cases} a_0 & \text{para } x > 15 \\ a_1 & \text{para } x \leqq 15 \end{cases}$

(c) $d_3(x) = \begin{cases} a_0 & \text{para } x > 16 \\ a_1 & \text{para } x \leqq 16? \end{cases}$

APLICACIONES

12.17 Una aerolínea quiere probar la hipótesis nula de que 60 por ciento de sus pasajeros objetan a que se fume dentro el avión. Explique en qué condiciones cometerían un error de tipo I y en qué condiciones cometerían un error de tipo II.

12.18 Se le pide a un doctor que haga un examen físico general muy completo a un ejecutivo para probar la hipótesis nula de que él podrá encargarse de responsabilidades adicionales. Explique en qué condiciones el doctor cometería un error de tipo I y en qué condiciones el doctor cometería un error de tipo II.

12.19 El tiempo promedio de secado de la pintura de una fabricante es 20 minutos. Investigue la efectividad de una modificación en la composición química de su pintura, la fabricante quiere probar la hipótesis nula $\mu = 20$ minutos contra una alternativa apropiada, donde μ es el tiempo promedio de secado de la pintura modificada.

(a) ¿Qué hipótesis alternativa debe usar la fabricante si no quiere hacer la modificación en la composición química de la pintura a menos que reduzca el tiempo de secado?

(b) ¿Qué hipótesis alternativa debe usar la fabricante si el nuevo proceso es realmente más barato y ella quiere hacer la modificación a menos que aumente el tiempo de secado de la pintura?

12.20 El departamento de policía de una ciudad está considerando reemplazar los neumáticos en sus autos con neumáticos radiales. Si μ_1 es el número promedio de millas que duran los neumáticos anteriores y μ_2 es el número promedio de millas que durarán los nuevos neumáticos, la hipótesis nula que debe probarse es $\mu_1 = \mu_2$.

(a) ¿Que hipótesis alternativa debe usar el departamento si no quiere usar los neumáticos radiales a menos que se pruebe definitivamente que dan un mejor millaje? En otras palabras, se pone el peso de la prueba en los neumáticos radiales, y se retendrán los neumáticos anteriores a menos que se rechace la hipótesis nula.

(b) ¿Que hipótesis alternativa debe usar el departamento si está ansioso en obtener los neumáticos radiales a menos que realmente den un millaje más pobre que los neumáticos anteriores? Advierta que el nuevo peso de la prueba está en los neumáticos anteriores, los cuales se retendrán sólo si se puede rechazar la hipótesis nula.

(c) ¿Que hipótesis alternativa debe usar el departamento de manera que el rechazo de la hipótesis nula pueda llevar a mantener los neumáticos anteriores o a comprar los nuevos?

12.21 Un botánico desea probar la hipótesis nula de que el diámetro promedio de las flores de una planta en particular es 9.6 cm. Decide tomar una muestra aleatoria de tamaño $n = 80$ y aceptar la hipótesis nula si la media de la muestra cae entre 9.3 cm y 9.9 cm; si la media de esta muestra cae fuera de este intervalo, él rechazará la hipótesis nula. ¿Qué decisión tomará y estará en error si

(a) obtiene una media de la muestra de 10.2 cm y $\mu = 9.6$ cm;

(b) obtiene una media de la muestra de 10.2 cm y $\mu = 9.8$ cm;

(c) obtiene una media de la muestra de 9.2 cm y $\mu = 9.6$ cm;

(d) obtiene una media de la muestra de 9.2 cm y $\mu = 9.8$ cm?

12.22 Un especialista en educación está considerando el uso de material de instrucción en audio casetes para una clase especial de estudiantes de tercer año con deficiencias en lectura. A los estudiantes en esta clase se les da una prueba estandarizada en mayo del año escolar, y μ_1 es la puntuación promedio obtenida en estas pruebas después de muchos años de experiencia. Sea μ_2 la puntuación promedio de los estudiantes que usan los audio casetes, y suponga que las puntuaciones altas son deseables.

(a) ¿Qué hipótesis nula debe usar el especialista en educación?

(b) ¿Qué hipótesis alternativa debe usarse si el especialista no desea adoptar los nuevos casetes a menos que mejoren las puntuaciones de la prueba estandarizada?

(c) ¿Qué hipótesis alternativa debe usarse si el especialista desea adoptar los nuevos casetes a menos que empeoren las puntuaciones de la prueba estandarizada?

12.23 Suponga que queremos probar la hipótesis nula de que un dispositivo anticontaminante para los autos es efectivo.

(a) Explique en qué condiciones cometeríamos un error de tipo I y bajo que condiciones cometeríamos un error de tipo II.

(b) El que un error sea un error de tipo I o un error de tipo II depende de cómo formulemos la hipótesis nula. Reexprese la hipótesis nula de manera que un error de tipo I se vuelva un error de tipo II, y viceversa.

12.24 Una bióloga quiere probar la hipótesis nula de que la envergadura media de cierta clase de insectos es 12.3 mm contra la hipótesis alternativa de que no es 12.3 mm. Si toma una muestra aleatoria y decide aceptar la hipótesis nula si y sólo si la media de la muestra cae entre 12.0 mm y 12.6 mm, ¿qué decisión tomará si obtiene $\bar{x} = 12.9$ mm y estará equivocada si

(a) $\mu = 12.5$ mm; (b) $\mu = 12.3$ mm?

12.25 Un empleado bancario quiere probar la hipótesis nula de que en promedio el banco paga 10 cheques malos por día contra la alternativa de que esta cifra es demasiado pequeña. Si toma una muestra aleatoria y decide rechazar la hipótesis nula si y sólo si la media de la muestra excede 12.5, ¿qué decisión tomará si obtiene $\bar{x} = 11.2$, y estará equivocado si

(a) $\lambda = 11.5$; (b) $\lambda = 10.0$?

En este caso λ es la media de la población de Poisson que se está muestreando.

12.26 Rehaga el ejemplo 12.3 con

(a) $\beta = 0.03$; (b) $\beta = 0.01$.

12.27 Suponga que queremos probar la hipótesis nula de que cierta clase de neumático durará, en promedio, 35,000 millas contra la hipótesis alterna de que durará, en promedio, 45,000 millas. Suponga que estamos tratando con una variable aleatoria que tiene una distribución exponencial, especificamos el tamaño de la muestra y la probabilidad de un error de tipo I y use el lema Neyman-Pearson para construir una región crítica. ¿Obtendríamos la misma región crítica si cambiamos la hipótesis alternativa a

(a) $\theta_1 = 50{,}000$ miles; (b) $\theta_1 > 35{,}000$ miles?

12.5 LA FUNCIÓN DE POTENCIA DE UNA PRUEBA

En el ejemplo 12.1 pudimos dar valores únicos a las probabilidades de cometer errores de tipo I y tipo de II porque estábamos probando una hipótesis simple contra una alternativa simple. En la práctica real, es relativamente raro, sin embargo, que las hipótesis simples se prueben contra alternativas simples; usualmente una o la otra, o ambas, son compuestas. Por ejemplo, en el ejemplo 12.1, bien podríamos haber sido mas realistas al probar la hipótesis nula que la tasa de recuperación de la enfermedad es $\theta \geqq 0.90$ contra la hipótesis alternativa $\theta < 0.90$, esto es, la hipótesis alternativa de que el nuevo medicamento no es tan efectivo como se afirma.

Cuando tratamos con hipótesis compuestas, el problema de evaluar los méritos de un criterio de prueba, o región crítica, se vuelve más complejo. En ese caso tenemos que considerar las probabilidades $\alpha(\theta)$ de cometer un error de tipo I para todos los valores de θ dentro del dominio especificado bajo la hipótesis nula H_0 y las probabilidades $\beta(\theta)$ de cometer un error de tipo II dentro de θ del dominio especificado bajo la hipótesis alternativa H_1. Se acostumbra combinar los dos conjuntos de probabilidades de la siguiente manera

DEFINICIÓN 12.3 La **función de potencia** de una prueba de una hipótesis estadística H_0 contra una hipótesis alternativa H_1 está dada por

$$\pi(\theta) = \begin{cases} \alpha(\theta) & \text{para los valores de } \theta \text{ asumidos bajo } H_0 \\ 1 - \beta(\theta) & \text{para los valores de } \theta \text{ asumidos bajo } H_1 \end{cases}$$

Así, los valores de la función de potencia son las probabilidades de rechazar la hipótesis nula H_0 para diversos valores del parámetro θ. Observe también que para los valores de θ asumidos bajo H_0, la función de potencia da la probabilidad de cometer un error de tipo I, y para los valores de θ asumidos bajo H_1, da la probabilidad de *no* cometer un error de tipo II.

EJEMPLO 12.5

Con respecto al ejemplo 12.1, suponga que hubiésemos querido probar la hipótesis nula $\theta \geqq 0.90$ contra la hipótesis alternativa $\theta < 0.90$. Investigue la función de potencia correspondiente al mismo criterio de prueba como en la página 386, donde aceptamos la hipótesis nula si $x > 14$ y la rechazamos si $x \leqq 14$. Como antes, x es el número observado de éxitos (recuperaciones) en $n = 20$ intentos.

Solución

Al escoger valores de θ para los cuales las probabilidades respectivas, $\alpha(\theta)$ o $\beta(\theta)$, están disponibles de la tabla I, encontramos las probabilidades $\alpha(\theta)$ de obtener cuando mucho 14 éxitos para $\theta = 0.90$ y 0.95, y las probabilidades $\beta(\theta)$ de obtepodrían usarse para probar un hipótesis nula dada contra una alternativa dada. Incidentalmente, si hubiésemos graficado en la figura 12.2 las probabilidades de aceptar H_0 (en vez de las de rechazar H_0), hubiésemos obtenido la **curva característica de operación**, o simplemente la **curva OC**, de la región crítica dada. En otras palabras, los valores de la función característica de operación, usados principalmente en aplicaciones industriales, están dados por $1 - \pi(\theta)$.

En la página 390 indicamos que en la teoría de Neyman-Pearson de prueba de hipótesis mantenemos fija α, la probabilidad de un error de tipo I, y esto requiere que la hipótesis nula H_0 sea una hipótesis simple, digamos, $\theta = \theta_0$. Como resultado, la función de potencia de cualquier prueba de esta hipótesis nula pasará por el punto (θ_0, α), el único punto en el cual el valor de una función de potencia es la probabilidad de cometer un error. Esto facilita la comparación de las funciones de potencia de varias regiones críticas, todas las cuales están diseñadas para probar la hipótesis nula simple $\theta = \theta_0$ contra una alternativa compuesta, digamos, la hipótesis alternativa $\theta \neq \theta_0$. Para ilustrar, considere la figura 12.3, que da las funciones de potencia de tres regiones críticas diferentes, o criterios de prueba, diseñadas para este propósito. Puesto que para cada valor de θ, excepto θ_0, los valores de las funciones de potencia son las probabilidades de tomar las decisiones correctas, es deseable tenerlas tan cercanas a 1 como sea posible. Así, se puede ver por inspección que la región crítica cuya función de potencia está dada por la curva punteada de la figura 12.3 es preferible a la región crítica cuya función de potencia está dada por la curva punteada. La probabilidad de no cometer un error de tipo II con la primera de estas regiones críticas siempre excede al de la segunda, y decimos que la primera región crítica es **uniformemente más potente** que la segunda; también se dice que la segunda región crítica es **inadmisible**.

La misma distinción clara no es posible si intentamos comparar las regiones críticas cuyas funciones de potencia estén dadas por las curvas punteadas y sólidas de la figura 12.3; en este caso es preferible la primera para $\theta < \theta_0$, mientras que la otra es preferible para $\theta > \theta_0$. En situaciones como ésta necesitamos criterios adicionales

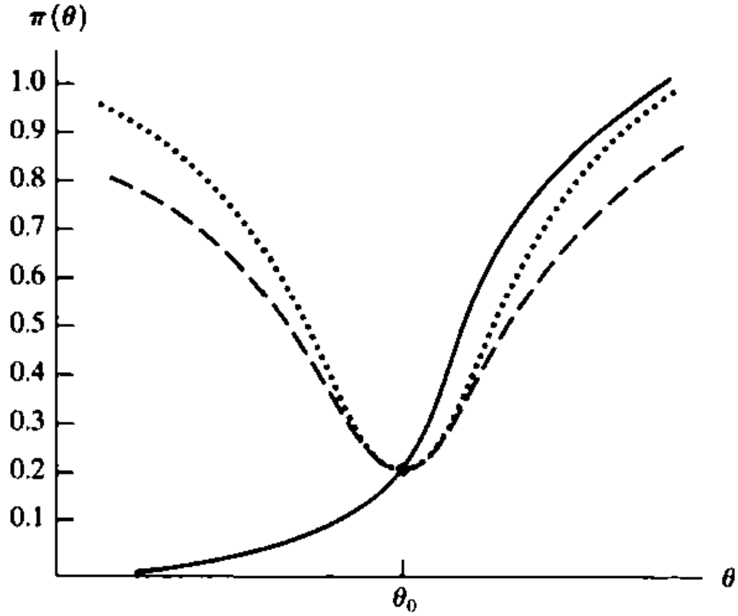

Figura 12.3 Funciones de potencia.

y a poblaciones continuas, pero todos nuestros argumentos se pueden extender fácilmente al caso multiparamétrico y a poblaciones discretas.

Para ilustrar la técnica de la razón de verosimilitud, supongamos que $X_1, X_2, \ldots, X_n$ constituyen una muestra aleatoria de tamaño n de una población cuya densidad en x es $f(x; \theta)$ y que Ω es el conjunto de valores que el parámetro θ puede asumir. A menudo nos referimos a Ω como **el espacio de parámetro** para θ. La hipótesis nula que queremos probar es

$$H_0: \quad \theta \in \omega$$

y la hipótesis alternativa es

$$H_1: \quad \theta \in \omega'$$

donde ω es un subconjunto de Ω y ω' es el complemento de ω con respecto a Ω. Así el espacio de parámetro para θ se divide en dos conjuntos ajenos ω y ω'; de acuerdo a la hipótesis nula, θ es un elemento del primer conjunto, y de acuerdo a la hipótesis alternativa es un elemento del segundo conjunto. En la mayoría de los problemas Ω es cualquiera de: el conjunto de todos los números reales, el conjunto de todos los números reales positivos, algún intervalo de números reales, o un conjunto discreto de números reales.

Cuando H_0 y H_1 son ambos hipótesis simples, ω y ω' tienen cada uno sólo un elemento, y en la sección 12.4 construimos pruebas para comparar las verosimilitudes L_0 y L_1. En el caso general, donde al menos una de las dos hipótesis es compuesta, comparamos en vez de ello las dos cantidades máx L_0 y máx L, donde máx L_0 es el valor máximo de la función de verosimilitud (véase página 346) para todos los valores de θ en ω, y máx L es el valor máximo de la función de verosimilitud para todos los valores de θ en Ω. En otras palabras, si tenemos una muestra aleatoria de tamaño n de una población cuya densidad en x es $f(x; \theta)$, $\hat{\theta}$ es la estimación de máxima verosimilitud de θ sujeta a la restricción que θ debe ser un elemento de ω, y $\hat{\hat{\theta}}$ es la estimación de máxima verosimilitud de θ para todos los valores de θ en Ω, entonces

$$\text{máx } L_0 = \prod_{i=1}^{n} f(x_i; \hat{\theta})$$

y

$$\text{máx } L = \prod_{i=1}^{n} f(x_i; \hat{\hat{\theta}})$$

Ambas cantidades son valores de variables aleatorias, puesto que dependen de los valores observados $x_1, x_2, \ldots, x_n$, y su razón

$$\lambda = \frac{\text{máx } L_0}{\text{máx } L}$$

se conoce como un valor de la **estadística de la razón de verosimilitud** Λ (griega mayúscula *lambda*).

Puesto que máx L_0 y máx L son ambas valores de una función de verosimilitud y por consiguiente nunca son negativas, se sigue que $\lambda \geqq 0$; también, puesto ω es un subconjunto del espacio de parámetro Ω, se sigue que $\lambda \leqq 1$. Cuando la hipótesis nula es falsa, esperaríamos que máx L_0 sea pequeña comparada con máx L, en cuyo caso λ sería cercana a cero. Por otra parte, cuando la hipótesis nula es verdadera y $\theta \in \omega$, esperaríamos que máx L_0 sea cercana a máx L, en cuyo caso λ sería cercana a 1. Una prueba de la razón de verosimilitud afirma, por consiguiente, que la hipótesis nula H_0 se rechace si y sólo si λ cae en la región crítica de la forma $\lambda \leqq k$, donde $0 < k < 1$. Para resumir:

DEFINICIÓN 12.4 Si ω y ω' son subconjuntos complementarios del espacio de parámetro Ω y si

$$\lambda = \frac{\text{máx } L_0}{\text{máx } L}$$

donde máx L_0 y máx L son los valores máximos de la función de verosimilitud para todos los valores de θ en ω y Ω, respectivamente, entonces la región crítica

$$\lambda \leqq k$$

donde $0 < k < 1$, define una **prueba de razón de** verosimilitud de la hipótesis nula $\theta \in \omega$ contra la hipótesis alternativa $\theta \in \omega'$.

Si H_0 es una hipótesis simple, se escoge k de manera que el tamaño de la región crítica sea igual a α; si H_0 es compuesta, k se escoge de manera que la probabilidad de un error de tipo I sea menor que, o igual a α para toda θ en ω, e igual a α, si es posible, para al menos un valor de θ en ω. Así, si H_0 es una hipótesis simple y $g(\lambda)$ es la densidad de Λ en λ cuando H_0 es verdad, entonces k debe ser tal que

$$P(\Lambda \leqq k) = \int_0^k g(\lambda)\,d\lambda = \alpha$$

En el caso discreto, la integral se reemplaza con una suma, y k se toma como el valor más grande para el cual la suma es menor que, o igual a α.

EJEMPLO 12.6

Encuentre la región crítica de la prueba de razón de verosimilitud para probar la hipótesis nula

$$H_0\colon \quad \mu = \mu_0$$

contra la alternativa compuesta

$$H_1\colon \quad \mu \neq \mu_0$$

sobre la base de una muestra aleatoria de tamaño n de una población normal con la varianza conocida σ^2.

Solución

Puesto que ω sólo contiene μ_0, se sigue que $\hat{\mu} = \mu_0$, y puesto que Ω es el conjunto de todos los números reales, se sigue por el método de la sección 10.7 que $\hat{\hat{\mu}} = \bar{x}$. Así:

En el ejemplo anterior fue fácil encontrar la constante que hizo el tamaño de la región crítica igual a α, porque pudimos referirnos a la distribución conocida de $\overline{X}$ y no tuvimos que derivar la distribución de la estadística de la razón de verosimilitud Λ misma. Puesto que la distribución de Λ es usualmente muy complicada, lo cual hace difícil evaluar k, a menudo es preferible usar la siguiente aproximación, al final del capítulo hacemos referencia a su prueba.

TEOREMA 12.2 Para n grande, la distribución de $-2 \cdot \ln \Lambda$ se aproxima, bajo condiciones muy generales, a la distribución ji cuadrada con 1 grado de libertad.

Debemos añadir que este teorema se aplica sólo al caso de un parámetro; si la población tiene más de un parámetro desconocido sobre los cuales la hipótesis nula impone r restricciones, el número de grados de libertad en la aproximación ji cuadrada a la distribución de $-2 \cdot \ln \Lambda$ es igual a r. Por ejemplo, si queremos probar la hipótesis nula de que la media y la varianza desconocidas de una población normal son μ_0 y σ_0^2 contra la hipótesis alternativa que $\mu \neq \mu_0$ y $\sigma^2 \neq \sigma_0^2$, el número de grados de libertad en la aproximación ji cuadrada a la distribución de $-2 \cdot \ln \Lambda$ sería 2; las dos restricciones son $\mu = \mu_0$ y $\sigma^2 = \sigma_0^2$.

Puesto que los valores pequeños de λ corresponden a valores grandes de $-2 \cdot \ln \lambda$, podemos usar el teorema 12.2 para escribir la región crítica de esta prueba aproximada de la razón de verosimilitud como

$$-2 \cdot \ln \lambda \geqq \chi^2_{\alpha, 1}$$

donde $\chi^2_{\alpha, 1}$ se define como en la página 282. En relación con el ejemplo 12.6 encontramos que

$$-2 \cdot \ln \lambda = \frac{n}{\sigma^2}(\bar{x} - \mu_0)^2 = \left(\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}\right)^2$$

que realmente *es* un valor de una variable aleatoria que tiene la distribución ji cuadrada con 1 grado de libertad.

Como indicamos en la página 400, la técnica de la razón de verosimilitud generalmente producirá resultados satisfactorios. Que éste no es siempre el caso se ilustra con el siguiente ejemplo, el cual es algo fuera de la común.

EJEMPLO 12.7

Sobre la base de una observación única, queremos probar la hipótesis nula simple que la distribución de probabilidad de X es

x	1	2	3	4	5	6	7
$f(x)$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

contra la alternativa compuesta que la distribución de probabilidad es

x	1	2	3	4	5	6	7
$g(x)$	$\frac{a}{3}$	$\frac{b}{3}$	$\frac{c}{3}$	$\frac{2}{3}$	0	0	0

donde $a + b + c = 1$. Demuestre que la región crítica obtenida por medio de la técnica de la razón de verosimilitud es inadmisible.

Solución

La hipótesis alternativa compuesta incluye todas las distribuciones de probabilidad que obtenemos al asignar valores diferentes de 0 a 1 a a, b y c, sujeto sólo a la restricción que $a + b + c = 1$. Para determinar λ para cada valor de x, primero hacemos $x = 1$. Para este valor obtenemos máx $L_0 = \frac{1}{12}$, máx $L = \frac{1}{3}$ (que corresponde a $a = 1$), y por tanto $\lambda = \frac{1}{4}$. Al determinar de la misma manera λ para los otros valores de x, obtenemos los resultados mostrados en la siguiente tabla:

x	1	2	3	4	5	6	7
λ	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{3}{8}$	1	1	1

Si el tamaño de la región crítica va a ser $\alpha = 0.25$, encontramos que la técnica de la razón de verosimilitud nos da una región crítica para la cual la hipótesis nula se rechaza cuando $\lambda = \frac{1}{4}$, esto es, cuando $x = 1$, $x = 2$ o $x = 3$; claramente, $f(1) + f(2) + f(3) = \frac{1}{12} + \frac{1}{12} + \frac{1}{12} = 0.25$. La probabilidad correspondiente de un error de tipo II está dada por $g(4) + g(5) + g(6) + g(7)$, y por tanto es igual a $\frac{2}{3}$.

Ahora consideremos la región crítica para la cual la hipótesis nula se rechaza sólo cuando $x = 4$. Su tamaño también es $\alpha = 0.25$ puesto que $f(4) = \frac{1}{4}$, pero la probabilidad correspondiente de un error de tipo II es

$$g(1) + g(2) + g(3) + g(5) + g(6) + g(7) = \frac{a}{3} + \frac{b}{3} + \frac{c}{3} + 0 + 0 + 0$$
$$= \frac{1}{3}$$

Puesto que esto es menos que $\frac{2}{3}$, la región crítica obtenida por medio de la técnica de la razón de verosimilitud es inadmisible. ▲

EJERCICIOS

12.28 Con respecto al ejercicio 12.3, suponga que hubiésemos querido probar la hipótesis nula $k \leqq 2$ contra la hipótesis alternativa $k > 2$. Encuentre las probabilidades de

(a) errores del tipo I para $k = 0$, 1 y 2;

(b) errores del tipo II para $k = 4$, 5, 6 y 7.

Dibuje también la gráfica de la función de potencia correspondiente.

12.29 Con respecto al ejemplo 12.5, suponga que rechazamos la hipótesis nula si $x \leqq 15$ y la aceptamos si $x > 15$. Calcule $\pi(\theta)$ para los mismos valores de θ

como en la tabla en la página 398 y dibuje la gráfica de la función de potencia de este criterio de prueba.

12.30 En la solución del ejemplo 12.6, verifique el paso que lleva a

$$\lambda = e^{-\frac{n}{2\sigma^2}(\bar{x}-\mu_0)^2}$$

12.31 El número de éxitos en n intentos se va a usar para probar la hipótesis nula de que el parámetro θ de una población binomial es igual a $\frac{1}{2}$ contra la alternativa que no es igual a $\frac{1}{2}$.

(a) Encuentre la expresión para la estadística de la razón de verosimilitud.

(b) Use el resultado del inciso (a) para demostrar que la región crítica de la prueba de razón de verosimilitud se puede escribir como

$$x \cdot \ln x + (n-x) \cdot \ln(n-x) \geqq K$$

donde x es el número observado de éxitos.

(c) Estudie la gráfica de $f(x) = x \cdot \ln x + (n-x) \cdot \ln(n-x)$, en particular su mínimo y su simetría, para demostrar que la región crítica de esta prueba de razón de verosimilitud también se puede escribir como

$$\left| x - \frac{n}{2} \right| \geqq K$$

donde K es una constante que depende del tamaño de la región crítica.

12.32 Una muestra aleatoria de tamaño n se va a usar para probar la hipótesis nula que el parámetro θ de una población exponencial es igual a θ_0 contra la alternativa de que no es igual a θ_0.

(a) Encuentre una expresión para la estadística de la razón de verosimilitud.

(b) Use el resultado del inciso (a) para demostrar que la región crítica de la prueba de razón de verosimilitud se puede escribir como

$$\bar{x} \cdot e^{-\bar{x}/\theta_0} \leqq K$$

12.33 Una muestra aleatoria de tamaño n de una población normal con media y varianza desconocidas se va a usar para probar la hipótesis nula $\mu = \mu_0$ contra la alternativa $\mu \neq \mu_0$. Use las estimaciones simultáneas de máxima verosimilitud de μ y σ^2 obtenidas en el ejemplo 10.17, demuestre que los valores de la estadística de la razón de verosimilitud se puede escribir en la forma

$$\lambda = \left(1 + \frac{t^2}{n-1}\right)^{-n/2}$$

donde $t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$. Note que la prueba de razón de verosimilitud puede así basarse en la distribución t.

12.34 Para la estadística de la razón de verosimilitud del ejercicio 12.33, muestre que $-2 \cdot \ln \lambda$ se aproxima a t^2 conforme $n \to \infty$. [*Sugerencia*: use la serie infinita para $\ln(1 + x)$ dada en la página 223.]

12.35 Dada una muestra aleatoria de tamaño n de una población normal con media y varianza desconocidas, encuentre una expresión para la estadística de la razón de verosimilitud para probar la hipótesis nula $\sigma = \sigma_0$ contra la hipótesis alternativa $\sigma \neq \sigma_0$. (*Sugerencia*: véase el ejemplo 10.17.)

12.36 Muestras aleatorias independientes de tamaño $n_1, n_2, \ldots,$ y n_k de k poblaciones normales con medias y varianzas desconocidas se van a usar para probar la hipótesis nula $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2$ contra la alternativa que estas varianzas no son todas iguales

(a) Demuestre que bajo la hipótesis nula las estimaciones de máxima verosimilitud de las medias μ_i y las varianzas σ_i^2 son

$$\hat{\mu}_i = \bar{x}_i \quad \text{y} \quad \hat{\sigma}_i^2 = \sum_{i=1}^{k} \frac{(n_i - 1)s_i^2}{n}$$

donde $n = \sum_{i=1}^{k} n_i$, mientras que sin restricciones las estimaciones de máxima verosimilitud de las medias μ_i y las varianzas σ_i^2 son

$$\hat{\hat{\mu}}_i = \bar{x}_i \quad \text{y} \quad \hat{\hat{\sigma}}_i^2 = \frac{(n_i - 1)s_i^2}{n_i}$$

Esto se sigue directamente de los resultados obtenidos en la sección 10.8.

(b) Use los resultados del inciso (a), demuestre que la estadística de la razón de verosimilitud se puede escribir como

$$\lambda = \frac{\prod_{i=1}^{k} \left[\frac{(n_i - 1)s_i^2}{n_i}\right]^{n_i/2}}{\left[\sum_{i=1}^{k} \frac{(n_i - 1)s_i^2}{n}\right]^{n/2}}$$

12.37 Muestre que para $k = 2$ la estadística de razón de verosimilitud del ejercicio 12.36 se puede expresar en términos de la razón de dos varianzas de muestra y que la prueba de razón de verosimilitud puede, por consiguiente, basarse en la distribución F.

12.38 Cuando probamos una hipótesis nula simple contra una alternativa compuesta, se dice que una región crítica es **insesgada** si la función de potencia correspondiente asume su valor mínimo en el valor del parámetro supuesto bajo la hipótesis nula. En otras palabras, una región crítica es insesgada si la probabilidad de rechazar la hipótesis nula es mínima cuando la hipótesis nula es verdadera. Dada una observación única de la variable aleatoria X que tiene la densidad

$$f(x) = \begin{cases} 1 + \theta^2\left(\frac{1}{2} - x\right) & \text{para } 0 < x < 1 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

Donde $-1 \leqq \theta \leqq 1$, muestre que la región crítica $x \leqq \alpha$ provee una región crítica insesgada de tamaño α para probar la hipótesis nula $\theta = 0$ contra la hipótesis alternativa $\theta \neq 0$.

APLICACIONES

12.39 Se va a usar una observación única para probar la hipótesis nula de que la media del tiempo de espera entre temblores registrados en una estación sismológica (la media de una población exponencial) es $\theta = 10$ horas contra la alternativa que $\theta \neq 10$ horas. Si la hipótesis nula será rechazada si y sólo si los valores observados son menos que 8 o mayores que 12, encuentre

(a) la probabilidad de un error de tipo I;

(b) las probabilidades de errores de tipo II cuando $\theta = 2, 4, 6, 8, 12, 16$ y 20.

Grafique también la función de potencia de este criterio de prueba.

12.40 Una muestra aleatoria de tamaño 64 se va a usar para probar la hipótesis nula de que para cierto grupo de edad la media de la puntuación en una prueba de rendimiento (la media de una población normal con $\sigma^2 = 256$) es menor que, o igual a, 40.0 contra la alternativa que es mayor que 40.0. Si la hipótesis nula se rechazará si y sólo si la media de la muestra aleatoria excede 43.5, encuentre

(a) las probabilidades de errores de tipo I cuando $\mu = 37.0, 38.0, 39.0$ y 40.0;

(b) las probabilidades de errores de tipo II cuando $\mu = 41.0, 42.0, 43.0, 44.0, 45.0, 46.0, 47.0$ y 48.0.

Grafique también la función de potencia de este criterio.

12.41 La suma de los valores obtenidos en una muestra aleatoria de tamaño $n = 5$ se va a usar para probar la hipótesis nula de que en promedio hay más de dos accidentes por semana en cierta intersección (que $\lambda > 2$ para esta población Poisson) contra la hipótesis alternativa de que en promedio el número de accidentes es de dos o menos. Si la hipótesis nula se rechazará si y sólo si la suma de las observaciones es cinco o menos, encuentre

(a) las probabilidades de errores de tipo I cuando $\lambda = 2.2, 2.4, 2.6, 2.8$ y 3.0;

(b) las probabilidades de errores de tipo II cuando $\lambda = 2.0, 1.5, 1.0$ y 0.5.

(*Sugerencia*: use los resultados del ejemplo 7.15.) También, dibuje la gráfica de la función de potencia de este criterio de prueba.

12.42 Verifique la afirmación de la página 400 de que 57 caras y 43 cruces en 100 lanzamientos de una moneda no nos permiten rechazar la hipótesis nula de que la moneda está perfectamente balanceada (contra la alternativa de que no está perfectamente balanceada) al nivel 0.05 de significancia. (*Sugerencia*: use la aproximación normal a la distribución binomial.)

12.43 Al comparar las variaciones en peso de cuatro razas de perros, los investigadores tomaron muestras aleatorias independientes de tamaño $n_1 = 8$, $n_2 = 10$, $n_3 = 6$ y $n_4 = 8$, y obtuvieron $s_1^2 = 16$, $s_2^2 = 25$, $s_3^2 = 12$ y $s_4^2 = 24$. Suponga que las poblaciones muestreadas son normales, use la fórmula del inciso (b) del ejercicio 12.36 para calcular $-2 \cdot \ln \lambda$ y pruebe la hipótesis nula $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2$ al nivel 0.05 de significancia. Explique por qué el número de grados de libertad para esta prueba aproximada de ji cuadrada es 3.

12.44 Los tiempos de avería de ciertos componentes electrónicos son 15, 28, 3, 12, 42, 19, 20, 2, 25, 30, 62, 12, 18, 16, 44, 65, 33, 51, 4 y 28 minutos. Si consideramos estos datos como una muestra aleatoria de una población exponencial, use los resultados de los ejercicios 12.32 y del teorema 12.2 para probar la hipótesis nula $\theta = 15$ minutos contra la hipótesis alternativa $\theta \neq 15$ minutos al nivel 0.05 de significancia. (Use $\ln 1.763 = 0.570$.)

REFERENCIAS

Los exámenes de diversas propiedades de las pruebas de razón de verosimilitud, particularmente sus propiedades de muestra grande, y una prueba del teorema 12.2 se pueden encontrar en la mayoría de los libros de texto avanzados sobre la teoría estadística, por ejemplo, en

LEHMANN, E. L., *Testing Statistical Hypotheses*. 2nd ed. Nueva York: John Wiley & Sons, Inc., 1986.

WILKS, S. S., *Mathematical Statistics*. Nueva York: John Wiley & Sons, Inc., 1962.

Mucho de la investigación original hecha en esta área se reproduce en

Selected Papers in Statistics and Probability by Abraham Wald. Stanford, Calif.: Stanford University Press, 1957.

CAPÍTULO

13

Prueba de hipótesis: aplicaciones

13.1 INTRODUCCIÓN

En el capítulo 12 examinamos una parte de la teoría que sustenta las pruebas estadísticas, y en este capítulo presentaremos algunas de las pruebas estándar que se usan más ampliamente en las aplicaciones. La mayoría de estas pruebas, al menos las que se basan en distribuciones conocidas de poblaciones, se puede obtener mediante la técnica de la razón de verosimilitud.

Para explicar la terminología que usaremos, consideremos una situación donde queremos probar la hipótesis nula H_0: $\theta = \theta_0$ contra la hipótesis **alternativa bilateral** H_1: $\theta \neq \theta_0$. Puesto que parece razonable aceptar la hipótesis nula cuando nuestra estimación puntual $\hat{\theta}$ de θ es cercana a θ_0 y rechazarla cuando $\hat{\theta}$ es mucho más grande, o mucho más pequeña que θ_0, sería lógico dejar que la región crítica consista en ambas colas de la distribución muestral de nuestra estadística $\hat{\Theta}$. Una prueba así se conoce como **prueba de dos colas**.

Por otra parte, si estamos probando la hipótesis nula H_0: $\theta = \theta_0$ contra la alternativa unilateral H_1: $\theta < \theta_0$, parecería razonable rechazar H_0 sólo cuando $\hat{\theta}$ es mucho más pequeña que θ_0. Por consiguiente, en este caso sería lógico dejar que la región crítica consista sólo en la cola del lado izquierdo de la distribución muestral de $\hat{\Theta}$. De igual manera, al probar H_0: $\theta = \theta_0$ contra la alternativa unilateral H_1: $\theta > \theta_0$, rechazamos H_0 sólo para valores grandes de $\hat{\theta}$, y la región crítica consiste solamente en la cola derecha de la distribución muestral de $\hat{\Theta}$. Cualquier prueba donde la región crítica consiste sólo en una cola de la distribución muestral de la estadística de prueba se llama **prueba de una cola**.

Por ejemplo, para la alternativa bilateral $\mu \neq \mu_0$ en el ejemplo 12.6, la técnica de la razón de verosimilitud lleva a una prueba de dos colas con la región crítica

$$|\bar{x} - \mu_0| \geqq z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

o

$$\bar{x} \leqq \mu_0 - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \quad \text{y} \quad \bar{x} \geqq \mu_0 + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

Como se dibuja en la figura 13.1, la hipótesis nula $\mu = \mu_0$ se rechaza si $\bar{X}$ asume un valor que caiga en cualquiera de las colas de su distribución muestral. En forma simbólica, esta región crítica se puede escribir como $z \leqq -z_{\alpha/2}$ o $z \geqq z_{\alpha/2}$, donde

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

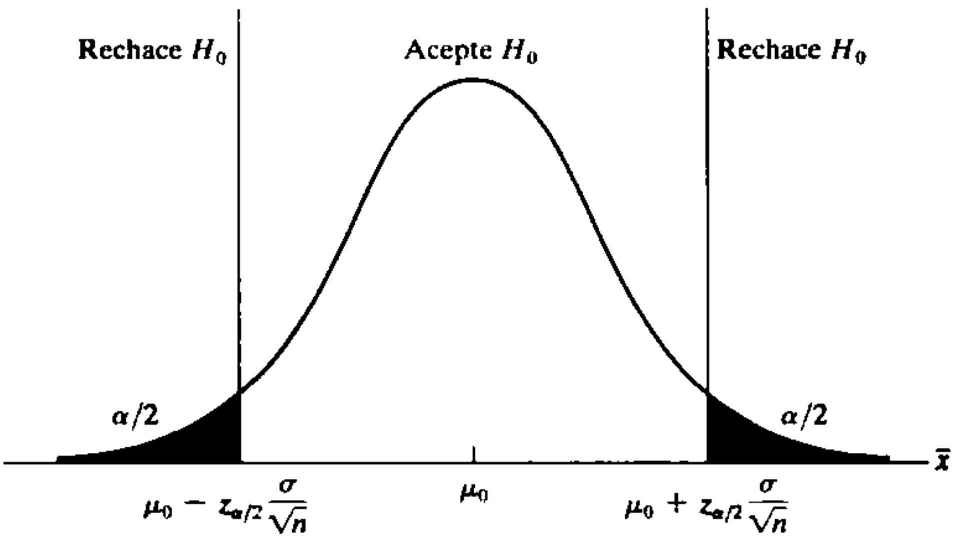

Figura 13.1 Región crítica para la prueba de dos colas.

Si hubiésemos usado la alternativa unilateral $\mu > \mu_0$, la técnica de la razón de verosimilitud nos hubiese llevado a la prueba de una cola cuya región crítica se dibuja en la figura 13.2, y si hubiésemos usado la alternativa unilateral $\mu < \mu_0$, la técnica de la razón de verosimilitud nos hubiese llevado a la prueba de una cola cuya región crítica se dibuja en la figura 13.3. Es lógico que en el primer caso rechazaríamos la hipótesis nula sólo para valores de $\bar{X}$ que caigan en la cola del lado derecho de su distribución muestral, y en el segundo caso rechazaríamos la hipótesis nula sólo para valores de $\bar{X}$ que caigan en la cola del lado izquierdo de su distribución muestral. De manera simbólica, las regiones críticas correspondientes se pueden escribir como $z \geqq z_\alpha$ y como $z \leqq -z_\alpha$, donde z se define como antes. Aunque hay excepciones a esta regla (véase el ejercicio 13.1), las alternativas bilaterales suelen llevar a pruebas de dos colas y las alternativas unilaterales suelen conducir hacia pruebas de una cola.

Tradicionalmente, ha sido costumbre bosquejar las pruebas de hipótesis mediante los siguientes pasos:

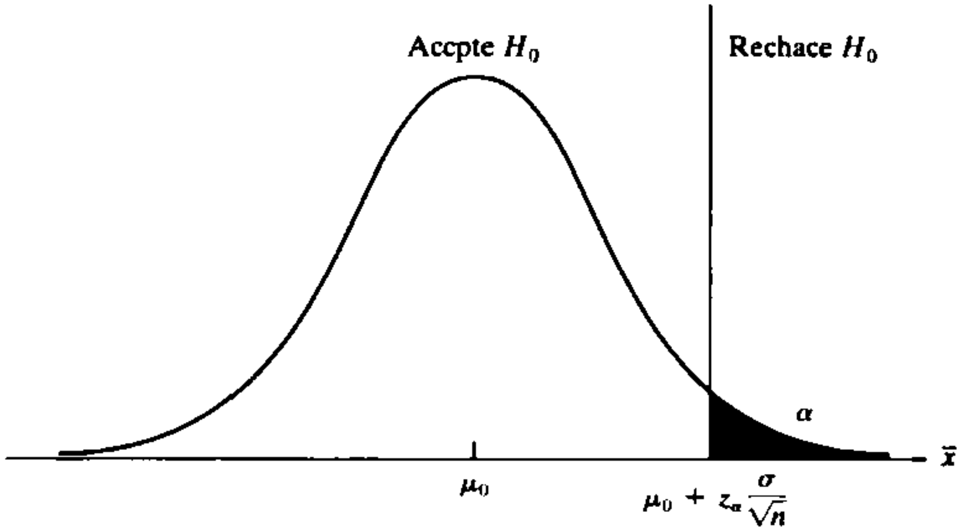

Figura 13.2 Región crítica para la prueba de una cola ($H_1: \mu > \mu_0$).

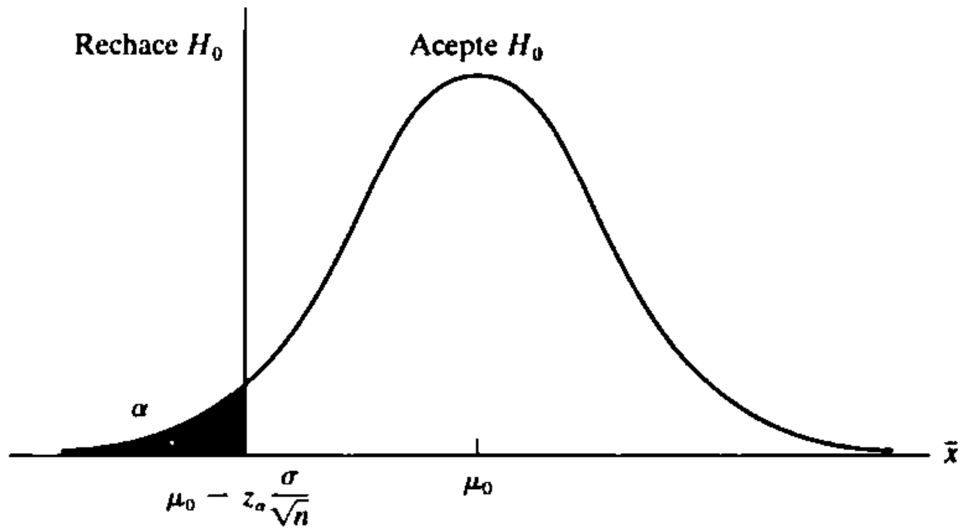

Figura 13.3 Región crítica para la prueba de una cola ($H_1: \mu < \mu_0$).

1. **Formule H_0 y H_1, y especifique α.**
2. **Use la distribución muestral de una estadística de prueba apropiada, determine una región crítica de tamaño α.**
3. **Determine el valor de la estadística de prueba a partir de los datos de la muestra.**
4. **Compruebe si el valor de la estadística de prueba cae en la región crítica y, con base en ello, rechace la hipótesis nula, o acéptela, o bien reserve el juicio.**

En las figuras 13.1, 13.2 y 13.3, las líneas divisorias del criterio de prueba (esto es, los **límites** de las regiones críticas, o los **valores críticos**) requieren el conocimiento de en la cola del lado derecho o en la cola del lado izquierdo de la distribución muestral de $\overline{X}$. En este caso se supone otra vez que la hipótesis nula es verdadera.

Más generalmente, definimos los valores P como sigue

> **DEFINICIÓN 13.1** El valor P es el menor nivel de significancia, que corresponde a un valor observado de la estadística de prueba, en el cual la hipótesis nula podría haberse rechazado.

Con respecto a este enfoque alternativo a probar hipótesis, el primero de los cuatro pasos en la página 412 permanece sin cambio, el segundo paso se convierte en

2′. Especifique la estadística de prueba.

y el tercer paso se convierte en

3′. Determine el valor de la estadística de prueba y el valor P correspondiente a partir de los datos de la muestra.

Y el cuarto paso se convierte en

4′. Compruebe **si el valor P es** menor **que, o igual a α y, de acuerdo a esto, rechace la hipótesis nula, o acéptela, o bien reserve el juicio.**

Como señalamos en la página 413, esto nos da más libertad en la selección del nivel de significancia, pero es difícil concebir situaciones donde podríamos justificar el uso de, digamos, $\alpha = 0.04$ en vez de $\alpha = 0.05$ o $\alpha = 0.015$ en vez de $\alpha = 0.01$. En la práctica, es prácticamente imposible evitar algún elemento de arbitrariedad, y en la mayoría de los casos hacemos juicios subjetivos, al menos en parte, si $\alpha = 0.05$ o $\alpha = 0.01$ refleja riesgos aceptables. Por supuesto, cuando hay mucho en juego y es práctico, podríamos usar un nivel de significancia mucho más pequeño que $\alpha = 0.01$.

En muchos casos, se debe entender que los dos métodos para probar hipótesis, los cuatro pasos que se dieron en la página 412, y los cuatro pasos aquí descritos, son equivalentes. Esto significa que sin importar el método que usemos, la decisión final (rechazar la hipótesis nula, aceptarla, o bien reservar el juicio) será la misma. En la práctica, usamos cualquier método que sea más conveniente, y esto puede depender de la distribución muestral de la estadística de prueba, la disponibilidad de tablas estadísticas o del software de computadora, y la naturaleza del problema (véase, por ejemplo, el ejemplo 13.8 y el ejercicio 13.52).

Hay estadísticos que prefieren evitar los problemas relacionados con la elección del nivel de significancia. Limitan su papel al análisis de datos, no especifican α y omiten el paso 4′. Por supuesto, siempre es deseable tener el punto de vista de otros (investigadores o gerentes) al formular la hipótesis y especificar α, pero difícilmente sería razonable dejar caer los valores P en manos de personas sin capacitación adecuada en estadística y dejar que ellos sigan a partir de ese punto. Para complicar las dificultades, considere la tentación a que uno podría estar expuesto al escoger α después de haber pastelería pierde dinero cuando $\mu > 8$ y el cliente pierde cuando $\mu < 8$, pruebe la hipótesis nula $\mu = 8$ contra la hipótesis alternativa $\mu \neq 8$ al nivel 0.01 de significancia.

Solución

1. H_0: $\mu = 8$
   H_1: $\mu \neq 8$
   $\alpha = 0.01$
2. Rechace la hipótesis nula si $z \leqq -2.575$ o $z \geqq 2.575$, donde
$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$
3. Al sustituir $\bar{x} = 8.091$, $\mu_0 = 8$, $\sigma = 0.16$ y $n = 25$, obtenemos
$$z = \frac{8.091 - 8}{0.16/\sqrt{25}} = 2.84$$
4. Puesto que $z = 2.84$ excede a 2.575, se debe rechazar la hipótesis nula y se deben hacer ajustes apropiados en el proceso de producción. ▲

Si hubiésemos usado el enfoque alternativo descrito en la página 414, hubiésemos obtenido un valor P de 0.0046 (véase el ejercicio 13.8), y puesto que 0.0046 es menos que 0.01, la conclusión hubiera sido la misma.

Se debe observar que la región crítica $z \geqq z_\alpha$ también se puede usar para probar la hipótesis nula $\mu = \mu_0$ contra la alternativa simple $\mu = \mu_1 > \mu_0$ o la hipótesis nula compuesta $\mu \leqq \mu_0$ contra la alternativa compuesta $\mu > \mu_0$. En el primer caso estaríamos probando una hipótesis simple contra una alternativa simple como en la sección 12.4 (véase el ejemplo 12.4, donde estudiamos esta prueba para $\sigma = 1$), y en el segundo caso α sería la probabilidad máxima de cometer un error de tipo I para cualquier valor de μ asumido bajo la hipótesis nula. Por supuesto, se aplicarían argumentos similares a la región crítica $z \leqq -z_\alpha$.

Cuando estamos tratando con una muestra grande de tamaño $n \geqq 30$ de una población que no necesita ser normal pero que tiene una varianza finita, podemos usar el teorema del límite central para justificar el uso de la prueba para poblaciones normales, y aun cuando σ^2 es desconocida podemos aproximar su valor con s^2 en el cálculo de la estadística de prueba. Para ilustrar el uso de una **prueba aproximada de muestra grande**, considere el siguiente ejemplo.

EJEMPLO 13.2

Suponga que 100 neumáticos que cierto fabricante produce duraron en promedio 21,819 millas con una desviación estándar de 1,295 millas. Pruebe la hipótesis nula $\mu = 22{,}000$ millas contra la hipótesis alternativa $\mu < 22{,}000$ millas en el nivel 0.05 de significancia.

Solución

1. H_0: $\mu = 22{,}000$
   H_1: $\mu < 22{,}000$
   $\alpha = 0.05$
2. Rechace la hipótesis nula si $z \leqq -1.645$, donde

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

3. Al sustituir $\bar{x} = 21{,}819$, $\mu_0 = 22{,}000$, $s = 1{,}295$ por σ y $n = 100$, obtenemos

$$z = \frac{21{,}819 - 22{,}000}{1{,}295/\sqrt{100}} = -1.40$$

4. Puesto que $z = -1.40$ es mayor que -1.645, no se puede rechazar la hipótesis nula; no hay evidencia real de que los neumáticos no son tan buenos como se supone bajo la hipótesis nula. ▲

Si hubiésemos usado el enfoque alternativo descrito en la página 414, habríamos obtenido un valor P de 0.0808 (véase el ejercicio 13.9), que excede a 0.05. Como debiera haberse esperado, la conclusión es la misma: no se puede rechazar la hipótesis nula.

Cuando $n < 30$ y σ^2 es desconocida, no se puede usar la prueba que hemos estado examinando en esta sección. Sin embargo, en el ejercicio 12.33 vimos que para muestras aleatorias de poblaciones normales, la técnica de la razón de verosimilitud nos da una prueba correspondiente basada en

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

que, de acuerdo al teorema 8.13, es un valor de una variable aleatoria que tiene la distribución t con $n - 1$ grados de libertad. Así, las regiones críticas de tamaño α para probar la hipótesis nula $\mu = \mu_0$ contra las alternativas $\mu \neq \mu_0$, $\mu > \mu_0$ o $\mu < \mu_0$, son, respectivamente, $|t| \geqq t_{\alpha/2, n-1}$, $t \geqq t_{\alpha, n-1}$ y $t \leqq -t_{\alpha, n-1}$. Advierta que los comentarios de la página 416 en relación con la hipótesis alternativa $\mu_1 > \mu_0$ y la prueba de la hipótesis nula $\mu \leqq \mu_0$ contra la alternativa $\mu > \mu_0$ también se aplican en este caso.

El siguiente ejemplo ilustra esta **prueba t de una muestra,** como suele llamarse.

EJEMPLO 13.3

Las especificaciones para cierta clase de cinta piden una media de la resistencia al rompimiento de 185 libras. Si cinco piezas seleccionadas aleatoriamente de diferentes rollos tienen una resistencia al rompimiento de 171.6, 191.8, 178.3, 184.9 y 189.1 libras, pruebe la hipótesis nula $\mu = 185$ libras contra la hipótesis alternativa $\mu < 185$ libras en el nivel 0.05 de significancia.

Si no se puede sostener la suposición de varianzas iguales en un problema de esta clase, hay varias posibilidades. Una relativamente simple consiste en formar aleatoriamente parejas de los valores en las dos muestras y después considerar sus diferencias como una muestra aleatoria de tamaño n_1 o n_2, la que sea más pequeña, de una población normal que, bajo la hipótesis nula, tiene la media $\mu = \delta$. Después probamos esta hipótesis nula contra la alternativa apropiada por medio de los métodos de la sección 13.2. Ésta es una buena razón para tener $n_1 = n_2$, pero existen técnicas alternativas para manejar el caso donde $n_1 \neq n_2$ (una de éstas, la prueba *Smith-Satterthwaite*, se menciona en las referencias al final del capítulo).

Hasta ahora hemos limitado nuestro examen a muestras aleatorias que son independientes, y los métodos que presentamos en esta sección no se pueden usar, por ejemplo, para decidir con base en los pesos "antes y después" si cierta dieta es realmente eficaz o si las diferencias observadas entre los IQ promedio de maridos y sus esposas son realmente significativas. En ambos ejemplos las muestras no son independientes porque los datos están realmente *asociados por parejas*. Una forma común de manejar esta clase de problemas es proceder como en el párrafo anterior, esto es, trabajar con las diferencias entre las parejas de medidas u observaciones. Si n es grande, podemos entonces usar la prueba descrita en la página 415 para probar la hipótesis nula $\mu_1 - \mu_2 = \delta$ contra la alternativa apropiada, y si n es pequeña, podemos usar la prueba t descrita en la página 417, siempre y cuando las diferencias se puedan considerar como una muestra aleatoria de una población normal.

EJERCICIOS

13.1 Dada una muestra aleatoria de tamaño n de una población normal con varianza conocida σ^2, demuestre que la hipótesis nula $\mu = \mu_0$ se puede probar contra la hipótesis alternativa $\mu \neq \mu_0$ con el uso de un criterio de una cola que se base en la distribución ji cuadrada.

13.2 Suponga que una muestra aleatoria de una población normal con la varianza conocida σ^2 se va a usar para probar la hipótesis nula $\mu = \mu_0$ contra la hipótesis alternativa $\mu = \mu_1$, donde $\mu_1 > \mu_0$, y que las probabilidades de errores de tipo I y tipo II van a tener los valores preasignados α y β. Demuestre que el tamaño requerido de la muestra está dado por

$$n = \frac{\sigma^2(z_\alpha + z_\beta)^2}{(\mu_1 - \mu_0)^2}$$

13.3 Con respecto al ejercicio anterior, encuentre el tamaño requerido de la muestra cuando $\sigma = 9$, $\mu_0 = 15$, $\mu_1 = 20$, $\alpha = 0.05$ y $\beta = 0.01$.

13.4 Suponga que muestras aleatorias de tamaño n de dos poblaciones normales con varianzas conocidas σ_1^2 y σ_2^2 se van a usar para probar la hipótesis nula $\mu_1 - \mu_2 = \delta$ contra la hipótesis alternativa $\mu_1 - \mu_2 = \delta'$ y que las probabilidades de errores de tipo I y tipo II van a tener los valores preasignados α y β. Demuestre que el tamaño requerido de la muestra está dado por

$$n = \frac{(\sigma_1^2 + \sigma_2^2)(z_\alpha + z_\beta)^2}{(\delta - \delta')^2}$$

13.16 En 12 corridas de prueba sobre una ruta señalada, una lancha motora recientemente diseñada promedió 33.6 segundos con una desviación estándar de 2.3 segundos. Suponiendo que es razonable tratar los datos como una muestra aleatoria de una población normal, use los cuatro pasos en la página 412 para probar la hipótesis nula $\mu = 35$ contra la alternativa $\mu < 35$ en el nivel 0.05 de significancia.

13.17 Cinco mediciones del contenido de alquitrán de cierta clase de cigarrillos dieron 14.5, 14.2, 14.4, 14.3 y 14.6 mg/cigarrillo. Suponga que los datos son una muestra aleatoria de una población normal, use los cuatro pasos de la página 412 para demostrar que en el nivel 0.05 de significancia se debe rechazar la hipótesis nula $\mu = 14.0$ en favor de la alternativa $\mu \neq 14.0$.

13.18 Con respecto al ejercicio 13.17, demuestre que si la primera medición se registra incorrectamente como 16.0 en vez de 14.5, esto invertiría el resultado. Explique la paradoja aparente que aunque ha aumentado la diferencia entre la media de la muestra y μ_0 ya no es significativa.

13.19 Con respecto al ejercicio 13.17, use software estadístico apropiado para encontrar el valor P que corresponde al valor observado de la estadística de prueba. Use este valor P para resolver de nuevo el ejercicio.

13.20 Si la misma hipótesis se prueba con bastante frecuencia, es probable que se rechace por lo menos una vez, aun si es verdad. Un profesor de biología, al intentar demostrar este hecho, hizo que ratones blancos recorrieran un laberinto para determinar si los ratones blancos recorrían el laberinto más rápido que la norma establecida en muchas pruebas anteriores que incluyeron ratones de diferentes colores.

(a) Si el profesor realiza este experimento una vez con varios ratones (y usa el nivel 0.05 de significancia), ¿cuál es el nivel de probabilidad de que obtendrá un resultado "significativo" aun si el color del ratón no afecta su velocidad al recorrer el laberinto?

(b) Si el profesor repite el experimento con un nuevo conjunto de ratones blancos, ¿cuál es la probabilidad de que al menos uno de los experimentos dará un resultado "significativo" aun si el color del ratón no afecta su velocidad al recorrer el laberinto?

(c) Si el profesor hace que 30 de sus estudiantes hagan el mismo experimento, cada uno con un grupo diferente de ratones blancos, ¿cuál es la probabilidad de que al menos uno de estos experimentos resultará "significativo" aun si el color del ratón no juega ningún papel en la velocidad al recorrer el laberinto?

13.21 Un epidemiólogo está tratando de descubrir la causa de cierta clase de cáncer. Estudia un grupo de 10,000 personas por cinco años, mide 48 "factores" diferentes entre los que estaban hábitos alimenticios, hábitos de beber bebidas alcohólicas, de fumar, de hacer ejercicio, y así sucesivamente. Su objetivo es determinar si hay alguna diferencia en las medias de estos factores (variables) entre quienes desarrollaron el cáncer dado y quienes no. Él supone que estas variables son independientes, aun cuando puede haber evidencia en contrario. En un esfuerzo por ser cautelosamente conservador, usa el nivel 0.01 de significancia en todas sus pruebas estadísticas.

13.29 Para comparar dos clases de protectores de defensas, se montaron seis de cada clase en cierta marca de auto compacto. Entonces cada auto se hizo chocar contra una pared de concreto a 5 millas por hora, y los siguientes son los costos de la reparaciones (en dólares):

Protector de defensas 1: 127 168 143 165 122 139

Protector de defensas 2: 154 135 132 171 153 149

Use los cuatro pasos de la página 412 para probar en el nivel 0.01 de significancia si la diferencia entre las medias de estas dos muestras es significativa.

13.30 Con respecto al ejercicio 13.29, use software estadístico apropiado para encontrar el valor P que corresponde al valor observado de la estadística de prueba. Use este valor P para rehacer el ejercicio.

13.31 En un estudio sobre la eficacia de ciertos ejercicios para reducir de peso, un grupo de 16 personas hicieron estos ejercicios durante un mes y mostraron los siguientes resultados

Peso antes	*Peso después*	*Peso antes*	*Peso después*
211	198	172	166
180	173	155	154
171	172	185	181
214	209	167	164
182	179	203	201
194	192	181	175
160	161	245	233
182	182	146	142

Use el nivel 0.05 de significancia para probar la hipótesis nula $\mu_1 - \mu_2 = 0$ contra la hipótesis alternativa $\mu_1 - \mu_2 > 0$, y así juzgar si los ejercicios son eficaces en la reducción de peso.

13.32 Las siguientes son las pérdidas semanales promedio de horas de trabajo a causa de accidentes en 10 plantas industriales antes y después de poner en operación cierto programa de seguridad:

45 y 36, 73 y 60, 46 y 44, 124 y 119, 33 y 35

57 y 51, 83 y 77, 34 y 29, 26 y 24, 17 y 11

Use los cuatro pasos en la página 412 y el nivel 0.05 de significancia para probar si el programa de seguridad es eficaz.

13.33 Con respecto al ejercicio 13.32, use software estadístico apropiado para encontrar el valor P que corresponda al valor observado de la estadística de prueba. Use este valor P para volver a resolver el ejercicio.

13.4 PRUEBAS CONCERNIENTES A VARIANZAS

Hay varias razones por las que es importante probar las hipótesis concernientes a las varianzas de las poblaciones. En lo que concierne a las aplicaciones directas, un fabricante que tiene que cumplir con especificaciones rígidas tendrá que efectuar pruebas sobre la variabilidad de su producto, tal vez un maestro desea saber si ciertas aseveraciones son verdaderas acerca de la variabilidad que puede esperar en el desempeño de un estudiante, y quizá un farmacéutico tiene que comprobar si la variación en la potencia de una medicina está dentro de los límites permisibles. En lo que concierne a aplicaciones indirectas, las pruebas acerca de las varianzas a menudo son prerrequisitos para las pruebas concernientes a otros parámetros. Por ejemplo, la prueba t de dos muestras descrita en la página 420 requiere que las varianzas de las dos poblaciones sean iguales, y en la práctica esto significa que quizá tengamos que comprobar la razonabilidad de esta suposición antes de efectuar la prueba concerniente a las medias.

Entre las pruebas que estudiaremos en esta sección está una prueba de la hipótesis nula de que la varianza de una población normal es igual a una constante dada y la prueba de razón de verosimilitud de la igualdad de las varianzas de dos poblaciones normales (a la que nos referimos en el ejercicio 13.27).

La primera de estas pruebas es esencialmente la del ejercicio 12.35. Dada una muestra aleatoria de tamaño n de una población normal, queremos probar la hipótesis nula $\sigma^2 = \sigma_0^2$ contra una de las alternativas $\sigma^2 \neq \sigma_0^2$, $\sigma^2 > \sigma_0^2$ o $\sigma^2 < \sigma_0^2$ y, como el lector debe haber descubierto en el ejercicio 12.35, la técnica de la razón de verosimilitud nos lleva a una prueba que se basa en s^2, el valor de la varianza de la muestra. Basado en el teorema 8.10, podemos escribir así las regiones críticas para probar la hipótesis nula contra las dos alternativas de un lado como $\chi^2 \geqq \chi^2_{\alpha, n-1}$ y $\chi^2 \leqq \chi^2_{1-\alpha, n-1}$, donde

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

En lo que concierne a la alternativa bilateral, rechazamos la hipótesis nula si $\chi^2 \geqq \chi^2_{\alpha/2, n-1}$ o $\chi^2 \leqq \chi^2_{1-\alpha/2, n-1}$, y el tamaño de todas estas regiones críticas es, por supuesto, igual a α.

EJEMPLO 13.6

Suponga que el espesor de una parte usada de un semiconductor es su dimensión crítica y que las mediciones del espesor de una muestra aleatoria de 18 de dichas partes tiene la varianza $s^2 = 0.68$, donde las mediciones son en milésimas de una pulgada. El proceso se considera que está bajo control si la variación del espesor está dada por una varianza no mayor que 0.36. Suponga que las mediciones constituyen una muestra aleatoria de una población normal, pruebe la hipótesis nula $\sigma^2 = 0.36$ contra la hipótesis alternativa $\sigma^2 > 0.36$ en el nivel 0.05 de significancia.

que si la variable aleatoria X tiene la distribución F con ν_1 y ν_2 grados de libertad, entonces $\frac{1}{X}$ tiene la distribución F con ν_2 y ν_1 grados de libertad.

EJEMPLO 13.7

Al comparar la variabilidad de la resistencia a la tracción de dos clases de acero estructural, un experimento dio los resultados siguientes: $n_1 = 13$, $s_1^2 = 19.2$, $n_2 = 16$ y $s_2^2 = 3.5$, donde las unidades de medición son 1,000 libras por pulgada cuadrada. Suponga que las mediciones constituyen variables aleatorias independientes de dos poblaciones normales, prueba la hipótesis nula $\sigma_1^2 = \sigma_2^2$ contra la alternativa $\sigma_1^2 \neq \sigma_2^2$ en el nivel 0.02 de significancia.

Solución

1. H_0: $\sigma_1^2 = \sigma_2^2$
   H_1: $\sigma_1^2 \neq \sigma_2^2$
   $\alpha = 0.02$
2. Puesto que $s_1^2 \geqq s_2^2$, rechace la hipótesis nula si $\frac{s_1^2}{s_2^2} \geqq 3.67$, donde 3.67 es el valor de $f_{0.01,\,12,\,15}$.
3. Al sustituir $s_1^2 = 19.2$ y $s_2^2 = 3.5$, obtenemos
$$\frac{s_1^2}{s_2^2} = \frac{19.2}{3.5} = 5.49$$
4. Puesto que $f = 5.49$ excede a 3.67, se debe rechazar la hipótesis nula; concluimos que la variabilidad de la resistencia a la tracción de las dos clases de acero no es la misma. ▲

EJERCICIOS

13.34 Al hacer uso del hecho que la distribución ji cuadrada se puede aproximar con una distribución normal cuando ν, el número de grados de libertad, es grande, demuestre que para muestras grandes de poblaciones normales
$$s^2 \geqq \sigma_0^2\left[1 + z_\alpha\sqrt{\frac{2}{n-1}}\right]$$
es una región crítica aproximada de tamaño α para probar la hipótesis nula $\sigma^2 = \sigma_0^2$ contra la alternativa $\sigma^2 > \sigma_0^2$. Construya también las regiones críticas correspondientes para probar esta hipótesis nula contra las alternativas $\sigma^2 < \sigma_0^2$ y $\sigma^2 \neq \sigma_0^2$ (véase el ejercicio 8.38).

13.35 Haga uso del resultado del ejercicio 8.43, demuestre que para muestras aleatorias grandes de poblaciones normales, las pruebas de la hipótesis nula $\sigma^2 = \sigma_0^2$ se puede basar en la estadística:

$$\left(\frac{s}{\sigma_0} - 1\right)\sqrt{2(n-1)}$$

que tiene aproximadamente la distribución normal estándar.

APLICACIONES

13.36 Nueve determinaciones del calor específico del hierro dieron una desviación estándar de 0.0086. Suponga que estas determinaciones constituyen una muestra aleatoria de una población normal, pruebe la hipótesis nula $\sigma = 0.0100$ contra la hipótesis alternativa $\sigma < 0.0100$ en el nivel 0.05 de significancia.

13.37 En una muestra aleatoria, los pesos de 24 reses Black Angus de cierta edad tienen una desviación estándar de 238 libras. Suponga que los pesos constituyen una muestra aleatoria de una población normal, pruebe la hipótesis nula $\sigma = 250$ libras contra la alternativa bilateral $\sigma \neq 250$ libras en el nivel 0.01 de significancia.

13.38 En una muestra aleatoria, $s = 2.53$ minutos para la cantidad de tiempo que 30 mujeres tardaron en terminar la prueba escrita para su licencia de conducir. En el nivel 0.05 de significancia, pruebe la hipótesis nula $\sigma = 2.85$ minutos contra la hipótesis alternativa $\sigma < 2.85$ minutos (use el método descrito en el texto).

13.39 Use el método del ejercicio 13.35 para rehacer el ejercicio 13.38.

13.40 Los datos pasados indican que la desviación estándar de mediciones que inspectores experimentados hicieron en estampados de plancha metálica es 0.41 pulgadas cuadradas. Si un inspector nuevo mide 50 estampados con una desviación estándar de 0.49 pulgadas cuadradas, use el método del ejercicio 13.35 para probar la hipótesis nula $\sigma = 0.41$ pulgadas cuadradas contra la hipótesis alternativa $\sigma > 0.41$ pulgadas cuadradas en el nivel 0.05 de significancia.

13.41 Con respecto al ejercicio 13.40, encuentre el valor P que corresponde al valor observado de la estadística de prueba y úselo para decidir si la hipótesis nula podría haberse rechazado en el nivel 0.015 de significancia.

13.42 Con respecto al ejemplo 13.5, pruebe la hipótesis nula $\sigma_1 - \sigma_2 = 0$ contra la hipótesis alternativa $\sigma_1 - \sigma_2 > 0$ en el nivel 0.05 de significancia.

13.43 Con respecto al ejercicio 13.27, pruebe en el nivel 0.10 de significancia si es razonable suponer que las dos poblaciones muestreadas tienen varianzas iguales.

13.44 Con respecto al ejercicio 13.29, pruebe en el nivel 0.02 de significancia si es razonable suponer que las dos poblaciones muestreadas tienen varianzas iguales.

13.5 PRUEBAS CONCERNIENTES A PROPORCIONES

Si el resultado de un experimento es el número de votos que un candidato recibe en una votación, el número de defectos encontrados en una pieza de tela, el número de niños que se ausentan de la escuela en un día dado,..., nos referimos a estos datos como **datos de conteo.** Los modelos apropiados para el análisis de los datos de conteo son la distribución binomial, la distribución de Poisson, la distribución multinomial, y algunas de las demás distribuciones discretas que estudiamos en el capítulo 5. En esta sección presentaremos una de las pruebas más comunes basada en datos de conteo, una prueba concerniente al parámetro θ de la distribución binomial. Así, podríamos probar con base en una muestra si la verdadera proporción de curaciones de cierta enfermedad es 0.90 o si la verdadera proporción de defectos que salen en una línea de ensamble es 0.02.

En el ejercicio 12.12 se pidió al lector que mostrara que la región crítica más potente para probar la hipótesis nula $\theta = \theta_0$ contra la hipótesis alternativa $\theta = \theta_1 < \theta_0$, donde θ es el parámetro de una población binomial, se basa en el valor de X, el número de "éxitos" obtenidos en n ensayos. Cuando se trata de alternativas compuestas, la técnica de la razón de verosimilitud también brinda pruebas basadas en el número de éxitos observados (como vimos en el ejercicio 12.31 para el caso especial donde $\theta_0 = \frac{1}{2}$). De hecho, si queremos probar la hipótesis nula $\theta = \theta_0$ contra la alternativa unilateral $\theta > \theta_0$, la región crítica de tamaño α del criterio de la razón de verosimilitud es

$$x \geqq k_\alpha$$

donde k_α es el entero más pequeño para el cual

$$\sum_{y=k_\alpha}^{n} b(y; n, \theta_0) \leqq \alpha$$

y $b(y; n, \theta_0)$ es la probabilidad de obtener y éxitos en n ensayos binomiales cuando $\theta = \theta_0$. Por tanto el tamaño de esta región crítica, así como de las que siguen, es tan cercana como es posible a α sin excederla.

La región crítica correspondiente para probar la hipótesis nula $\theta = \theta_0$ contra la alternativa unilateral $\theta < \theta_0$ es

$$x \leqq k'_\alpha$$

donde k'_α es el entero más grande para el cual

$$\sum_{y=0}^{k'_\alpha} b(y; n, \theta_0) \leqq \alpha$$

y, finalmente, la región crítica para probar la hipótesis nula $\theta = \theta_0$ contra la alternativa bilateral $\theta \neq \theta_0$ es

$$x \geqq k_{\alpha/2} \quad \text{o} \quad x \leqq k'_{\alpha/2}$$

No ilustraremos este método de determinar las regiones críticas para pruebas concernientes al parámetro binomial θ porque, en la práctica real, es mucho menos tedioso basar las decisiones en los valores P.

EJEMPLO 13.8

Si $x = 4$ de $n = 20$ pacientes sufrieron efectos secundarios serios a causa de un nuevo medicamento, pruebe la hipótesis nula $\theta = 0.50$ contra la hipótesis alternativa $\theta \neq 0.50$ en el nivel 0.05 de significancia. En este caso θ es la proporción verdadera de pacientes que sufren efectos secundarios serios a causa del nuevo medicamento.

Solución

1. H_0: $\theta = 0.50$
   H_1: $\theta \neq 0.50$
   $\alpha = 0.05$

2'. Use la estadística de prueba X, el número observado de éxitos.

3'. $x = 4$, y puesto que $P(X \leqq 4) = 0.0059$, el valor P es $2(0.0059) = 0.0118$.

4'. Puesto que el valor P, 0.0118, es menor que 0.05, se debe rechazar la hipótesis nula; concluimos que $\theta \neq 0.50$. ▲

Las pruebas que hemos descrito requieren el uso de una tabla de probabilidades binomiales, sin importar si usamos los cuatro pasos de la página 412 o los de la página 414. Para $n \leqq 20$ podemos usar la tabla I al final del libro, y para valores de n hasta 100 podemos usar las tablas a las que hicimos referencia al final del capítulo 5. Alternativamente, para valores grandes de n podemos usar la aproximación normal de la distribución binomial y tratar

$$z = \frac{x - n\theta}{\sqrt{n\theta(1 - \theta)}}$$

como una variable aleatoria que tiene la distribución normal estándar. Para n grande, podemos probar así la hipótesis nula $\theta = \theta_0$ contra las alternativas $\theta \neq \theta_0$, $\theta > \theta_0$ o $\theta < \theta_0$ al usar, respectivamente, las regiones críticas $|z| \geqq z_{\alpha/2}$, $z \geqq z_\alpha$ y $z \leqq -z_\alpha$, donde

$$z = \frac{x - n\theta_0}{\sqrt{n\theta_0(1 - \theta_0)}}$$

o

$$z = \frac{\left(x \pm \frac{1}{2}\right) - n\theta_0}{\sqrt{n\theta_0(1 - \theta_0)}}$$

si usamos la corrección por continuidad introducida en el ejemplo 6.5. Usamos el signo menos cuando x excede a $n\theta_0$ y el signo de más cuando x es menor que $n\theta_0$.

EJEMPLO 13.9

Una compañía petrolera afirma que menos del 20 por ciento de los propietarios de autos no han probado su gasolina. Pruebe esta afirmación en el nivel 0.01 de significancia si una comprobación aleatoria revela que 22 de 200 propietarios de autos no han probado la gasolina de la compañía.

Solución

1. H_0: $\theta = 0.20$
   H_1: $\theta < 0.20$
   $\alpha = 0.01$
2. Rechace la hipótesis nula de $z \leq -2.33$, donde (sin la corrección por continuidad)

$$z = \frac{x - n\theta_0}{\sqrt{n\theta_0(1 - \theta_0)}}$$

3. Al sustituir $x = 22$, $n = 200$ y $\theta_0 = 0.20$, obtenemos

$$z = \frac{22 - 200(0.20)}{\sqrt{200(0.20)(0.80)}} = -3.18$$

4. Puesto que $z = -3.18$ es menos que -2.33, se debe rechazar la hipótesis nula; concluimos que, como se afirma, menos del 20% de todos los propietarios de autos no han probado la gasolina de la compañía. ▲

Advierta que si hubiésemos usado la corrección por continuidad en el ejemplo anterior, habríamos obtenido $z = -3.09$ y la conclusión habría sido la misma.

13.6 PRUEBAS CONCERNIENTES A DIFERENCIAS ENTRE *k* PROPORCIONES

En muchos problemas de investigación aplicada, debemos decidir si las diferencias observadas entre proporciones muestrales, o los porcentajes, son significativos o si se pueden atribuir a la suerte. Por ejemplo, si el 6 por ciento de los pollos congelados en la muestra de un proveedor falla en cumplir ciertos estándares y sólo 4 por ciento en la muestra de otro proveedor falla en cumplir los estándares, quizá deseamos investigar si la diferencia entre estos dos porcentajes es significativa. De la misma manera, tal vez deseamos juzgar sobre la base de datos muestrales si proporciones iguales de votantes en cuatro ciudades diferentes favorecen a cierto candidato a gobernador.

Para indicar un método general de manejar los problemas de esta clase, suponga que $x_1, x_2, \ldots, x_k$ son los valores observados de k variables aleatorias independientes $X_1, X_2, \ldots, X_k$ que tienen distribuciones binomiales con los parámetros n_1 y θ_1, n_2 y $\theta_2, \ldots, n_k$ y θ_k. Si las n son suficientemente grandes, podemos aproximar las distribuciones de las variables aleatorias independientes:

$$Z_i = \frac{X_i - n_i\theta_i}{\sqrt{n_i\theta_i(1-\theta_i)}} \qquad \text{para } i = 1, 2, \ldots, k$$

con distribuciones normales estándar, y, de acuerdo al teorema 8.8, podemos entonces considerar

$$\chi^2 = \sum_{i=1}^{k} \frac{(x_i - n_i\theta_i)^2}{n_i\theta_i(1-\theta_i)}$$

como un valor de una variable aleatoria que tiene la distribución ji cuadrada con k grados de libertad. Para probar la hipótesis nula, $\theta_1 = \theta_2 = \cdots = \theta_k = \theta_0$ (contra la alternativa que al menos una de las θ no es igual a θ_0), podemos usar así la región crítica $\chi^2 \geqq \chi^2_{\alpha,k}$, donde

$$\chi^2 = \sum_{i=1}^{k} \frac{(x_i - n_i\theta_0)^2}{n_i\theta_0(1-\theta_0)}$$

Cuando no se especifica θ_0 esto es, cuando sólo nos interesa la hipótesis nula $\theta_1 = \theta_2 = \cdots = \theta_k$, sustituimos por θ la estimación ponderada

$$\hat{\theta} = \frac{x_1 + x_2 + \cdots + x_k}{n_1 + n_2 + \cdots + n_k}$$

y la región crítica se vuelve $\chi^2 \geqq \chi^2_{\alpha,k-1}$, donde

$$\chi^2 = \sum_{i=1}^{k} \frac{(x_i - n_i\hat{\theta})^2}{n_i\hat{\theta}(1-\hat{\theta})}$$

La pérdida de 1 grado de libertad, esto es, el cambio de la región crítica de $\chi^2_{\alpha,k}$ a $\chi^2_{\alpha,k-1}$, se debe al hecho que se sustituye una estimación por el parámetro desconocido θ; en una referencia de la página 448 se hace un examen formal de esto.

Presentemos ahora una fórmula alternativa para la estadística ji cuadrada inmediata anterior, la cual, como veremos en la sección 13.7, se presta más rápidamente a otras aplicaciones. Si arreglamos los datos como en la tabla siguiente

	Éxitos	*Fracasos*
Muestra 1	x_1	$n_1 - x_1$
Muestra 2	x_2	$n_2 - x_2$
	$\cdots$	$\cdots$
Muestra k	x_k	$n_k - x_k$

las frecuencias de celda esperadas son

$$e_{11} = 400(0.53) = 212 \quad \text{y} \quad e_{12} = 400(0.47) = 188$$
$$e_{21} = 500(0.53) = 265 \quad \text{y} \quad e_{22} = 500(0.47) = 235$$
$$e_{31} = 400(0.53) = 212 \quad \text{y} \quad e_{32} = 400(0.47) = 188$$

y la sustitución en la fórmula para χ^2 dada arriba nos da

$$\begin{aligned}\chi^2 &= \frac{(232 - 212)^2}{212} + \frac{(260 - 265)^2}{265} + \frac{(197 - 212)^2}{212} \\ &\quad + \frac{(168 - 188)^2}{188} + \frac{(240 - 235)^2}{235} + \frac{(203 - 188)^2}{188} \\ &= 6.48\end{aligned}$$

4. Puesto que $\chi^2 = 6.48$ excede a 5.991, se debe rechazar la hipótesis nula; en otras palabras, las proporciones verdaderas de compradores que favorecen el detergente A sobre el detergente B en las tres ciudades no son las mismas. ▲

EJERCICIOS

13.45 Muestre que las dos fórmulas para χ^2 en la página 434 son equivalentes.

13.46 Modifique las regiones críticas en la página 430 de manera que se puedan usar para probar la hipótesis nula $\lambda = \lambda_0$ contra la hipótesis alternativa $\lambda > \lambda_0$, $\lambda < \lambda_0$ y $\lambda \neq \lambda_0$ sobre la base de n observaciones. En este caso λ es el parámetro de la distribución de Poisson. (*Sugerencia*: use el resultado del ejemplo 7.15.)

13.47 Con respecto al ejercicio 13.46, use la tabla II para encontrar los valores que corresponden a $k_{0.025}$ y $k'_{0.025}$ para probar la hipótesis nula $\lambda = 3.6$ contra la hipótesis alternativa $\lambda \neq 3.6$ sobre la base de cinco observaciones. Use el nivel 0.05 de significancia.

13.48 Parar $k = 2$, demuestre que la fórmula de χ^2 en la página 434 se puede escribir como

$$\chi^2 = \frac{(n_1 + n_2)(n_2x_1 - n_1x_2)^2}{n_1n_2(x_1 + x_2)[(n_1 + n_2) - (x_1 + x_2)]}$$

13.49 Dadas muestras aleatorias grandes de dos poblaciones binomiales, muestre que la hipótesis nula $\theta_1 = \theta_2$ se puede probar con base en la estadística

$$z = \frac{\dfrac{x_1}{n_1} - \dfrac{x_2}{n_2}}{\sqrt{\hat{\theta}(1 - \hat{\theta})\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$$

donde $\hat{\theta} = \dfrac{x_1 + x_2}{n_1 + n_2}$. (*Sugerencia*: refiérase al ejercicio 8.5.)

13.50 Muestre que el cuadrado de la expresión para z es igual a

$$\chi^2 = \sum_{i=1}^{2} \frac{(x_i - n_i\hat{\theta})^2}{n_i\hat{\theta}(1 - \hat{\theta})}$$

de manera que las dos pruebas son realmente equivalentes cuando la hipótesis alternativa es $\theta_1 \neq \theta_2$. Observe que la prueba descrita en el ejercicio 13.49, mas no la que se basa en la estadística χ^2 se puede usar cuando la hipótesis alternativa es $\theta_1 < \theta_2$ o $\theta_1 > \theta_2$.

APLICACIONES

13.51 Con respecto al ejemplo 13.8, muestre que la región crítica es $x \leqq 5$ o $x \geqq 15$ y que, correspondiendo a esta región crítica, el nivel de significancia es realmente 0.0414.

13.52 Se ha afirmado que más del 40 por ciento de todos los compradores pueden identificar una marca registrada a la que se le hace mucha publicidad. Si, en una muestra aleatoria, 10 de 18 compradores pudieron identificar la marca registrada, pruebe en el nivel 0.05 de significancia si la hipótesis nula $\theta = 0.40$ se puede rechazar contra la alternativa $\theta > 0.40$.

13.53 Con respecto al ejercicio 13.52, encuentre la región crítica y el nivel real de significancia que corresponde a esta región crítica.

13.54 Un doctor afirma que menos de 30 por ciento de todas las personas expuestas a cierta cantidad de radiación sentirán algún efecto dañino. Si, en una muestra aleatoria, sólo 1 de 19 personas expuestas a tal radiación sintió algún efecto dañino, prueba la hipótesis nula $\theta = 0.30$ contra la hipótesis alternativa $\theta < 0.30$ en el nivel 0.05 de significancia.

13.55 Con respecto al ejercicio 13.54, encuentre la región crítica y el nivel real de significancia que corresponde a esta región crítica.

13.56 En una muestra aleatoria, 12 de 14 accidentes industriales fueron causados por condiciones inseguras de trabajo. Use el nivel 0.01 de significancia para probar la hipótesis nula $\theta = 0.40$ contra la hipótesis alternativa $\theta \neq 0.40$.

13.57 Con respecto al ejercicio 13.56, encuentre la región crítica y el nivel real de significancia que corresponde a esta región crítica.

13.58 En una encuesta aleatoria de 1,000 hogares en Estados Unidos, se encontró que 29 por ciento de los hogares tenían al menos un miembro con un título universitario. ¿Este hallazgo refuta la aseveración de que la proporción de que todos estos hogares en Estados Unidos es al menos 35 por ciento? (Use el nivel 0.05 de significancia.)

13.59 En una encuesta aleatoria de 12 estudiantes no graduados de carreras comerciales, seis dijeron que tomarían cursos avanzados en contabilidad. Use el nivel 0.01 de significancia para probar la hipótesis nula $\theta = 0.20$, esto es, 20 por ciento de todos los estudiantes no graduados de carreras comerciales tomarán cursos avanzados de contabilidad, contra la hipótesis alternativa $\theta > 0.20$.

13.60 Un procesador de alimentos quiere saber si la probabilidad de que un cliente preferirá una nueva clase de empaque a la clase anterior es realmente 0.60. Si, en una muestra aleatoria, siete de 18 clientes prefieren la nueva clase de empa-

En este caso hay una muestra de tamaño 400, y los totales de los renglones así como los totales de las columnas se dejan al azar. Es principalmente en relación con problemas como éste que las tablas $r \times c$ se conocen como **tablas de contingencia**.

La hipótesis nula que queremos probar por medio de la tabla anterior es que el desempeño en el trabajo de las personas que han pasado por el programa de capacitación es independiente de su IQ. En general, si θ_{ij} es la probabilidad de que un elemento caerá en la celda que pertenece al *i*ésimo renglón y la *j*ésima columna, $\theta_{i\cdot}$ es la probabilidad de que un elemento caerá en el *i*esimo renglón, y $\theta_{\cdot j}$ es la probabilidad de que un elemento caerá en la *j*ésima columna, la hipótesis nula que queremos probar es

$$\theta_{ij} = \theta_{i\cdot} \cdot \theta_{\cdot j}$$

para $i = 1, 2, \ldots, r$ y $j = 1, 2, \ldots, c$. Correspondientemente, la hipótesis alternativa es $\theta_{ij} \neq \theta_{i\cdot} \cdot \theta_{\cdot j}$ para al menos un par de valores de i y j.

Puesto que el método por el cual analizamos una tabla $r \times c$ es el mismo sin importar si estamos tratando con r muestras de poblaciones multinomiales con c resultados diferentes o una muestra de una población multinomial con rc resultados diferentes, examinémoslo aquí con respecto al último. En el ejercicio 13.71 se pedirá al lector igualar el trabajo para la primera clase de problema.

En lo que sigue, denotaremos la frecuencia observada en el *i*ésimo renglón y la *j*ésima columna con f_{ij}, los totales de los renglones con $f_{i\cdot}$, los totales de las columnas con $f_{\cdot j}$, y el gran total, la suma de todas las frecuencias de las celdas, con f. Con esta notación, estimamos las probabilidades $\theta_{i\cdot}$ y $\theta_{\cdot j}$ como

$$\hat{\theta}_{i\cdot} = \frac{f_{i\cdot}}{f} \quad \text{y} \quad \hat{\theta}_{\cdot j} = \frac{f_{\cdot j}}{f}$$

y bajo la hipótesis nula de independencia obtenemos

$$e_{ij} = \hat{\theta}_{i\cdot} \cdot \hat{\theta}_{\cdot j} \cdot f = \frac{f_{i\cdot}}{f} \cdot \frac{f_{\cdot j}}{f} \cdot f = \frac{f_{i\cdot} \cdot f_{\cdot j}}{f}$$

para la frecuencia esperada para la celda en el *i*ésimo renglón y la *j*ésima columna. Advierta que e_{ij} así obtenida al ***multiplicar el total del renglón al cual pertenece la celda por el total de la columna a la cual pertenece y después dividir entre el gran total.***

Una vez que hemos calculado la e_{ij}, basamos nuestra decisión en el valor de

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(f_{ij} - e_{ij})^2}{e_{ij}}$$

y rechazamos la hipótesis nula si excede a $\chi^2_{\alpha,(r-1)(c-1)}$.

El número de grados de libertad es $(r-1)(c-1)$, y en relación con esto hagamos la siguiente observación: siempre que se estimen frecuencias de celdas en fórmulas de ji cuadrada con base en datos de conteo muestrales, el número de grados de libertad es $s - t - 1$, donde s es el número de términos en la suma y t es el número de parámetros independientes reemplazados por estimadores. Al hacer la prueba para las diferencias entre k proporciones con la estadística ji cuadrada de la sección 13.6, teníamos $s = 2k$ y $t = k$, puesto que teníamos que estimar los k parámetros θ_1, $\theta_2, \ldots, \theta_k$ y el número de grados de libertad fue $2k - k - 1 = k - 1$. Cuando hacemos la prueba para independencia en una tabla de contingencia $r \times c$ tenemos $s = rc$

y $t = r + c - 2$, puesto que los r parámetros $\theta_{i\cdot}$ y los c parámetros $\theta_{\cdot j}$ no son todos independientes: sus sumas respectivas deben ser igual a 1. Así, obtenemos $s - t - 1 = rc - (r + c - 2) - 1 = (r - 1)(c - 1)$.

Puesto que la estadística de prueba que hemos descrito sólo tiene aproximadamente una distribución ji cuadrada con $(r - 1)(c - 1)$ grados de libertad, es costumbre usar esta prueba sólo cuando ninguna de las e_{ij} es menor que 5; esto algunas veces requiere que combinemos algunas de las celdas con una pérdida correspondiente en el número de grados de libertad.

EJEMPLO 13.11

Use los datos mostrados en la siguiente tabla para probar en el nivel 0.01 de significancia si la habilidad de una persona en matemáticas es independiente de su interés en la estadística.

		Habilidad en matemáticas		
		Baja	*Promedio*	*Alta*
	Bajo	63	42	15
Interés en la estadística	*Promedio*	58	61	31
	Alto	14	47	29

Solución

1. H_0: La habilidad en matemáticas y el interés en la estadística son independientes.
   H_1: La habilidad en matemáticas y el interés en la estadística no son independientes.
   $\alpha = 0.01$
2. Rechace la hipótesis nula si $\chi^2 \geqq 13.277$, donde

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(f_{ij} - e_{ij})^2}{e_{ij}}$$

   y 13.277 es el valor de $\chi^2_{0.01,4}$.
3. Las frecuencias esperadas del primer renglón son $\frac{120 \cdot 135}{360} = 45.0$, $\frac{120 \cdot 150}{360} = 50.0$, y $120 - 45.0 - 50.0 = 25.0$, donde hicimos uso del hecho que para cada renglón o columna la suma de las frecuencias de celdas esperadas es igual a la suma de las frecuencias correspondientes observadas (véase el ejercicio 13.70). En forma similar, las frecuencias esperadas del segundo renglón son 56.25, 62.5 y 31.25, y las del tercer renglón (todas se obtuvieron al restar de los totales de las columnas) son 33.75, 37.5 y 18.75. Entonces, al sustituir en la fórmula para χ^2 nos da:

$$\chi^2 = \frac{(63 - 45.0)^2}{45.0} + \frac{(42 - 50.0)^2}{50.0} + \cdots + \frac{(29 - 18.75)^2}{18.75}$$

$$= 32.14$$

4. Puesto que $\chi^2 = 32.14$ excede a 13.277, se debe rechazar la hipótesis nula; concluimos que hay una relación entre la habilidad de una persona en matemáticas y su interés en la estadística. ▲

Una deficiencia del análisis ji cuadrada de una tabla $r \times c$ es que no toma en consideración un posible orden de los renglones y/o columnas. Por ejemplo, en el ejemplo 13.11, la habilidad en matemáticas así como el interés en la estadística se ordenan de bajo promedio a alto, y el valor que obtenemos para χ^2 permanecería igual si los renglones y/o las columnas se intercambiaran entre sí. También, las columnas de la tabla en la página 438 reflejan un orden de preferir B (no preferir A) a ser indiferentes a preferir A, pero en este caso no hay un orden específico de los renglones. La forma en que se puede tomar en consideración tal orden se explica en los ejercicios 14.61 y 15.12.

13.8 BONDAD DEL AJUSTE

La prueba de bondad del ajuste considerada aquí se aplica a situaciones en las que queremos determinar si un conjunto de datos se puede considerar como una muestra aleatoria de una población que tiene una distribución dada. En el capítulo 14 se examinará un segundo tipo de "bondad de ajuste" que se aplica al ajuste de una curva a un conjunto de pares de datos. Para ilustrar, suponga que queremos decidir, con base en los datos (frecuencias observadas) de la siguiente tabla, si el número de errores que un cajista hace al componer una galera de tipos es una variable aleatoria que tiene una distribución de Poisson:

Número de errores	*Frecuencia observada* f_i	*Probabilidades de Poisson con* $\lambda = 3$	*Frecuencias esperadas* e_i
0	18	0.0498	21.9
1	53	0.1494	65.7
2	103	0.2240	98.6
3	107	0.2240	98.6
4	82	0.1680	73.9
5	46	0.1008	44.4
6	18	0.0504	22.2
7	10	0.0216	9.5
8	2 } 3	0.0081	3.6 } 5.3
9	1	0.0038	1.7

Para determinar un conjunto correspondiente de frecuencias esperadas para una muestra aleatoria de una población de Poisson, primero usamos la media de la distribución observada para estimar el parámetro de Poisson λ, obtenemos $\hat{\lambda} = \frac{1,341}{440} = 3.05$ o, aproximadamente, $\hat{\lambda} = 3$. Después, copiamos las probabilidades de Poisson para $\lambda = 3$ de la tabla II (usamos la probabilidad de 9 *o más* en vez de la probabilidad de 9) y multiplicamos por 440, la frecuencia total, y obtenemos las frecuencias esperadas mostradas en la columna del lado derecho de la tabla. Para probar la hipótesis nula que las frecuencias observadas constituyen una muestra aleatoria de una población de Poisson, debemos juzgar qué tan buen ajuste tenemos, o qué tan próxima es la correlación, entre los dos conjuntos de frecuencias. En general, para probar la hipótesis nula H_0 que un conjunto de datos observados viene de una población que tiene una distribución especificada contra la alternativa de que la población tiene alguna otra distribución, calculamos

$$\chi^2 = \sum_{i=1}^{m} \frac{(f_i - e_i)^2}{e_i}$$

y rechazamos H_0 en el nivel α de significancia si $\chi^2 \geqq \chi^2_{\alpha, m-t-1}$, donde m es el número de términos en la suma y t es el número de parámetros independientes estimados con base en los datos muestrales (véase el análisis en las páginas 439 y 440). En el ejemplo anterior, $t = 1$ puesto que sólo se estima un parámetro con base en los datos, y el número de grados de libertad es $m - 2$.

EJEMPLO 13.12

Para los datos en la tabla 441, pruebe al nivel 0.05 de significancia si el número de errores que el cajista hace al componer una galera de tipos es una variable aleatoria que tiene una distribución de Poisson.

Solución

(Puesto que las frecuencias esperadas correspondientes a ocho y nueve errores son menores que 5, se combinan las dos clases.)

1. H_0: El número de errores es una variable aleatoria de Poisson.
   H_1: El número de errores no es una variable aleatoria de Poisson.
   $\alpha = 0.05$
2. Rechace la hipótesis nula si $\chi^2 \geqq 14.067$, donde
   $$\chi^2 = \sum_{i=1}^{m} \frac{(f_i - e_i)^2}{e_i}$$
   y 14.067 es el valor de $\chi^2_{0.05, 7}$.
3. Al sustituir en la fórmula para χ^2, obtenemos
   $$\chi^2 = \frac{(18 - 21.9)^2}{21.9} + \frac{(53 - 65.7)^2}{65.7} + \cdots + \frac{(3 - 5.3)^2}{5.3}$$
   $$= 6.83$$

4. Puesto que $\chi^2 = 6.83$ es menos que 14.067, no se puede rechazar la hipótesis nula; ciertamente, la proximidad de la correlación entre las frecuencias observadas y esperadas sugiere que la distribución de Poisson proporciona un "buen ajuste". ▲

EJERCICIOS

13.70 Verifique que las frecuencias de celda esperadas se calculan de acuerdo a la regla de la página 439, su suma para cualquier renglón o columna es igual a la suma de frecuencias observadas correspondientes.

13.71 Demuestre que la regla de la página 439 para calcular las frecuencias de celda esperadas también se aplica cuando probamos la hipótesis nula que estamos muestreando r poblaciones con distribuciones multinomiales idénticas.

13.72 Demuestre que la siguiente fórmula de cálculo para χ^2 es equivalente a la fórmula en la página 439:

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{f_{ij}^2}{e_{ij}} - f$$

13.73 Use la fórmula del ejercicio 13.72 para volver a calcular χ^2 para el ejemplo 13.10.

13.74 Si el análisis de una tabla de contingencia muestra que hay una relación entre las dos variables bajo consideración, la fortaleza de esta relación se puede medir con el **coeficiente de contingencia**

$$C = \sqrt{\frac{\chi^2}{\chi^2 + f}}$$

donde χ^2 es el valor obtenido para la estadística de prueba, y f es el gran total como se definió en la página 439. Demuestre que

(a) para una tabla de contingencia 2×2 el valor máximo de C es $\frac{1}{2}\sqrt{2}$;

(b) para una tabla de contingencia 3×3 el valor máximo de C es $\frac{1}{3}\sqrt{6}$.

APLICACIONES

13.75 Las muestras de un material experimental se producen mediante tres diferentes prototipos de procesos y se les hace una prueba de conformidad con un estándar de resistencia. Si las pruebas mostraron los resultados siguientes, ¿se puede decir en el nivel 0.01 de significancia que los tres procesos tienen la misma probabilidad de aprobar con este estándar de resistencia?

	Proceso A	*Proceso B*	*Proceso C*
Número que pasa la prueba	45	58	49
Número que falla la prueba	21	15	35

13.76 En un estudio sobre las opiniones de los padres de familia acerca de un curso obligatorio de educación sexual, 360 padres de familia, una muestra aleatoria, se clasificaron de acuerdo a si tienen uno, dos, tres o más hijos en el sistema escolar y también si opinan que el curso es malo, adecuado o bueno. Con base en los resultados que se muestran en la tabla siguiente, pruebe al nivel 0.05 de significancia si hay una relación entre la reacción al curso de los padres de familia y el número de hijos que tienen en el sistema escolar:

	Número de niños		
	1	2	3 *o más*
Malo	48	40	12
Adecuado	55	53	29
Bueno	57	46	20

13.77 Pruebas sobre la fidelidad y la selectividad de 190 radios produjeron los resultados que se muestran en la tabla siguiente:

		Fidelidad		
		Baja	*Promedio*	*Alta*
	Baja	7	12	31
Selectividad	*Promedio*	35	59	18
	Alta	15	13	0

Use el nivel 0.01 de significancia para probar la hipótesis nula de que la fidelidad es independiente de la selectividad.

13.78 Los siguientes datos muestrales corresponden a los embarques que recibió una empresa grande de tres proveedores diferentes

	Número de rechazados	*Número de imperfectos pero aceptables*	*Número de perfectos*
Proveedor A	12	23	89
Proveedor B	8	12	62
Proveedor C	21	30	119

Pruebe en el nivel 0.01 de significancia si los tres proveedores embarcan productos de igual calidad.

13.79 Analice la tabla de 3×3 de la página 438, que corresponde a las respuestas de compradores en tres ciudades diferentes con respecto a dos detergentes. Use el nivel 0.05 de significancia.

13.80 Se lanzaron cuatro monedas 160 veces y salieron 0, 1, 2, 3 o 4 caras, respectivamente, 19, 54, 58, 23 y 6 veces. Use el nivel 0.05 de significancia para probar si

(a) Verifique que la media y la desviación estándar de esta distribución son $\bar{x} = 20$ y $s = 5$.

(b) Encuentre las probabilidades de que una variable aleatoria que tiene la distribución normal con $\mu = 20$ y $\sigma = 5$ asumirá un valor menor que 9.5, entre 9.5 y 14.5, entre 14.5 y 19.5, entre 19.5 y 24.5, entre 24.5 y 29.5, entre 29.5 y 34.5, y mayor que 34.5.

(c) Encuentre la curva normal de frecuencias esperada para las diversas clases al multiplicar las probabilidades obtenidas en el inciso (b) por la frecuencia total, y después pruebe en el nivel 0.05 de significancia si se pueden considerar los datos como una muestra aleatoria de una población normal.

13.9 USO DE COMPUTADORAS

Al igual que en el capítulo 11, existe software de computadoras para todas las pruebas que hemos examinado. Una vez más, sólo tenemos que introducir los datos originales (sin tratar) en nuestra computadora junto con la instrucción apropiada. Para ilustrar, considere el ejemplo siguiente.

EJEMPLO 13.13

Las muestras aleatorias siguientes son mediciones de la capacidad calorífica (en millones de calorías por tonelada) de muestras de carbón de dos minas:

Mina 1:	8,400	8,230	8,380	7,860	7,930
Mina 2:	7,510	7,690	7,720	8,070	7,660

Use el nivel 0.05 de significancia para probar si la diferencia entre las medias de las dos muestras es significativa.

Solución

La impresión de computadora en la figura 13.5 muestra que el valor de la estadística de prueba es $t = 2.95$, el número de grados de libertad es 8, y el valor P es

```
MTB  > SET C1
DATA > 8400  8230  8380  7860  7930
MTB  > SET C2
DATA > 7510  7690  7720  8070  7660
MTB     > POOL C1 C2

TWOSAMPLE T FOR C1 VS C2
    N       MEAN     STDEV   SE MEAN
C1  5       8160       252       113
C2  5       7730       207        92

95 PCT CI FOR MU C1 - MU C2: (94, 766)
TTEST MU C1 = MU C2 (VS NE): [redacted]
```

Figura 13.5 Impresión de computadora para el ejemplo 13.13.

CAPÍTULO

14

Regresión y correlación

14.1 INTRODUCCIÓN

Un objetivo importante de muchas investigaciones estadísticas es establecer las relaciones que hagan posible predecir una o más variables en términos de otras. Así, se realizan estudios para predecir las ventas potenciales de un producto nuevo en términos de su precio, el peso de un paciente en términos del número de semanas que ha seguido un régimen alimenticio, los gastos familiares en entretenimiento en términos del ingreso familiar, el consumo *per cápita* de ciertos alimentos en términos de sus valores nutricionales y la cantidad de dinero que se gasta en hacerles publicidad en televisión, y así sucesivamente.

Aunque, por supuesto, es deseable poder predecir una cantidad exactamente en términos de otras, rara vez es posible, y en la mayoría de los casos tenemos que conformarnos con predecir promedios o valores esperados. Así, quizá no podamos predecir exactamente cuánto dinero ganará el Sr. Brown 10 años después de graduarse de la universidad; pero, dados los datos apropiados, podemos predecir el ingreso promedio de los graduados universitarios en términos del número de años transcurridos después de haber salido de la universidad. De la misma manera, en el mejor de los casos podemos predecir el rendimiento promedio de una variedad dada de trigo en términos de la precipitación pluvial en julio, y en el mejor de los casos podemos predecir el desempeño promedio de los estudiantes que inician estudios universitarios en términos de sus IQ.

Formalmente, si se nos da la distribución conjunta de dos variables aleatorias X y Y, y se sabe que X asume el valor x, el problema básico de la **regresión bivariada** es determinar la media condicional $\mu_{Y|x}$, esto es, el valor "promedio" de Y para el valor dado de X. El término "regresión", como se usa aquí, se remonta a Francis Galton, quien lo utilizó para indicar ciertas relaciones en la teoría de la herencia. En problemas que contienen más de dos variables aleatorias, esto es, en la **regresión múltiple**, tratamos con cantidades como $\mu_{Z|x,y}$, la media de Z para valores dados de X y Y, $\mu_{X_4|x_1,x_2,x_3}$, la media de X_4 para valores dados de X_1, X_2 y X_3, y así sucesivamente.

Si $f(x, y)$ es el valor de la densidad conjunta de dos variables aleatorias X y Y en (x, y), el problema de regresión bivariada es simplemente determinar la densidad condicional de Y dado $X = x$ y después evaluar la integral

$$\mu_{Y|x} = E(Y|x) = \int_{-\infty}^{\infty} y \cdot w(y|x)\, dy$$

como se bosquejó en la sección 4.8. La ecuación resultante se llama **ecuación de regresión** de Y sobre X. Alternativamente, tal vez nos interese la ecuación de regresión

$$\mu_{X|y} = E(X|y) = \int_{-\infty}^{\infty} x \cdot f(x|y)\, dx$$

En el caso discreto, cuando tratamos con distribuciones de probabilidad en vez de densidades de probabilidad, las integrales en las dos ecuaciones de regresión dadas arriba simplemente se reemplaza con sumas.

Cuando no conocemos la densidad de probabilidad conjunta o distribución de las dos variables aleatorias, o al menos no todos sus parámetros, la determinación de $\mu_{Y|x}$ o $\mu_{X|y}$ se vuelve un problema de estimación basado en datos muestrales; éste es un problema totalmente diferente, que examinaremos en las secciones 14.3 y 14.4.

EJEMPLO 14.1

Dadas las dos variables aleatorias X y Y que tienen la densidad conjunta

$$f(x, y) = \begin{cases} x \cdot e^{-x(1+y)} & \text{para } x > 0 \text{ y } y > 0 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

encuentre la ecuación de regresión de Y sobre X y bosqueje la curva de regresión

Solución

Al eliminar y por integración, encontramos que la densidad marginal de X está dada por

$$g(x) = \begin{cases} e^{-x} & \text{para } x > 0 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

y por tanto la densidad condicional de Y dado $X = x$ está dada por

$$w(y|x) = \frac{f(x, y)}{g(x)} = \frac{x \cdot e^{-x(1+y)}}{e^{-x}} = x \cdot e^{-xy}$$

para $y > 0$ y $w(y|x) = 0$ en cualquier otra parte, que reconocemos como una densidad exponencial con $\theta = \frac{1}{x}$. Por tanto, al evaluar

$$\mu_{Y|x} = \int_{0}^{\infty} y \cdot x \cdot e^{-xy}\, dy$$

o al referirnos al corolario 1 del teorema 6.3, encontramos que la ecuación de regresión de Y sobre X está dada por:

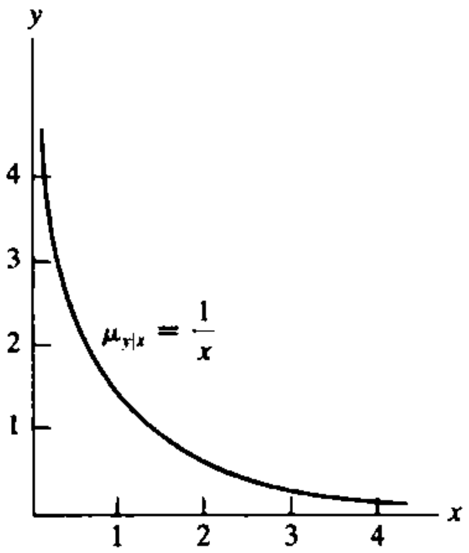

Figura 14.1 Curva de regresión del ejemplo 14.1.

$$\mu_{Y|x} = \frac{1}{x}$$

La curva de regresión correspondiente se muestra en la figura 14.1. ▲

EJEMPLO 14.2

Si X y Y tienen la distribución multinomial

$$f(x, y) = \binom{n}{x, y, n - x - y} \cdot \theta_1^x \theta_2^y (1 - \theta_1 - \theta_2)^{n-x-y}$$

para $x = 0, 1, 2, \ldots, n$ y $y = 0, 1, 2, \ldots, n$, con $x + y \leqq n$, encuentre la ecuación de regresión de Y sobre X.

Solución

La distribución marginal de X está dada por

$$\begin{aligned} g(x) &= \sum_{y=0}^{n-x} \binom{n}{x, y, n - x - y} \cdot \theta_1^x \theta_2^y (1 - \theta_1 - \theta_2)^{n-x-y} \\ &= \binom{n}{x} \theta_1^x (1 - \theta_1)^{n-x} \end{aligned}$$

para $x = 0, 1, 2, \ldots, n$, que reconocemos como una distribución binomial con los parámetros n y θ_1. Por tanto,

$$w(y|x) = \frac{f(x, y)}{g(x)} = \frac{\binom{n - x}{y} \theta_2^y (1 - \theta_1 - \theta_2)^{n-x-y}}{(1 - \theta_1)^{n-x}}$$

para $y = 0, 1, 2, \ldots, n - x$, y, al reescribir esta fórmula como:

$$w(y|x) = \binom{n-x}{y}\left(\frac{\theta_2}{1-\theta_1}\right)^y\left(\frac{1-\theta_1-\theta_2}{1-\theta_1}\right)^{n-x-y}$$

encontramos por inspección que la distribución condicional de Y dado $X = x$ es una distribución binomial con los parámetros $n - x$ y $\dfrac{\theta_2}{1-\theta_1}$, de manera que la ecuación de regresión de Y sobre X es

$$\mu_{Y|x} = \frac{(n-x)\theta_2}{1-\theta_1}$$

de acuerdo al teorema 5.2. ▲

Con respecto al ejemplo anterior, sea X el número de veces que sale un número par en 30 tiros de un dado balanceado y sea Y el número de veces que el resultado es cinco, entonces la ecuación de regresión se vuelve

$$\mu_{Y|x} = \frac{(30-x)\frac{1}{6}}{1-\frac{1}{2}} = \frac{1}{3}(30-x)$$

Esto es lógico, porque hay tres posibilidades igualmente probables, 1, 3 o 5, para cada uno de los $30 - x$ resultados que no son pares.

EJEMPLO 14.3

Si la densidad conjunta de X_1, X_2 y X_3 está dada por

$$f(x_1, x_2, x_3) = \begin{cases} (x_1 + x_2)e^{-x_3} & \text{para } 0 < x_1 < 1, 0 < x_2 < 1, x_3 > 0 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

encuentre la ecuación de regresión de X_2 sobre X_1 y X_3.

Solución

Al referirnos al ejemplo 3.22, encontramos que la densidad marginal de X_1 y X_3 está dada por

$$m(x_1, x_3) = \begin{cases} \left(x_1 + \frac{1}{2}\right)e^{-x_3} & \text{para } 0 < x_1 < 1, x_3 > 0 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

Por consiguiente,

$$\mu_{X_2|x_1,x_3} = \int_{-\infty}^{\infty} x_2 \cdot \frac{f(x_1, x_2, x_3)}{m(x_1, x_3)}\,dx_2 = \int_0^1 \frac{x_2(x_1 + x_2)}{\left(x_1 + \frac{1}{2}\right)}\,dx_2$$

$$= \frac{x_1 + \frac{2}{3}}{2x_1 + 1} \qquad ▲$$

Advierta que la esperanza condicional obtenida en el ejemplo anterior depende de x_1 pero no de x_3. Esto se podía haber esperado, puesto que indicamos en la página 123 que hay una independencia por parejas entre X_2 y X_3.

14.2 REGRESIÓN LINEAL

Una característica importante del ejemplo 14.2 es que la ecuación de regresión es **lineal**; esto es, es de la forma

$$\mu_{Y|x} = \alpha + \beta x$$

donde α y β son constantes, llamadas los **coeficientes de regresión**. Hay varias razones de por qué las ecuaciones de regresión lineal son de especial interés: primero, se prestan rápidamente a un tratamiento matemático adicional; después, a menudo proveen buenas aproximaciones a ecuaciones de regresión de otra forma complicadas; y finalmente, en el caso de la distribución normal bivariada, que estudiamos en la sección 6.7, las ecuaciones de regresión son, de hecho, lineales.

Para simplificar el estudio de las ecuaciones de regresión lineales, expresemos los coeficientes de regresión α y β en términos de algunos de los momentos más pequeños de la distribución conjunta de X y Y, esto es, en términos de $E(X) = \mu_1$, $E(Y) = \mu_2$, $\mathrm{var}(X) = \sigma_1^2$, $\mathrm{var}(Y) = \sigma_2^2$ y $\mathrm{cov}(X, Y) = \sigma_{12}$. Entonces, al usar también el coeficiente de correlación

$$\rho = \frac{\sigma_{12}}{\sigma_1 \sigma_2}$$

definido en la sección 6.7, podemos probar los siguientes resultados

> **TEOREMA 14.1** Si la regresión de Y sobre X es lineal, entonces
>
> $$\mu_{Y|x} = \mu_2 + \rho \frac{\sigma_2}{\sigma_1}(x - \mu_1)$$
>
> y si la regresión de X sobre Y es lineal, entonces
>
> $$\mu_{X|y} = \mu_1 + \rho \frac{\sigma_1}{\sigma_2}(y - \mu_2)$$

Demostración. Puesto que $\mu_{Y|x} = \alpha + \beta x$, se sigue que

$$\int y \cdot w(y|x)\, dy = \alpha + \beta x$$

y si multiplicamos la expresión en ambos lados de esta ecuación por $g(x)$, el valor correspondiente de la densidad marginal de X, e integramos sobre x, obtenemos

$$\int\int y \cdot w(y|x) g(x)\, dy\, dx = \alpha \int g(x)\, dx + \beta \int x \cdot g(x)\, dx$$

14.3 EL MÉTODO DE LOS MÍNIMOS CUADRADOS

En las secciones anteriores hemos examinado el problema de regresión sólo en relación con variables aleatorias que tienen distribuciones conjuntas. En la práctica real, hay muchos problemas donde un conjunto de **datos asociados en parejas** dan una indicación de que la regresión es lineal, donde no conocemos la distribución conjunta de las variables aleatorias en consideración pero, sin embargo, queremos estimar los coeficientes de regresión α y β. Los problemas de esta clase usualmente se manejan por el **método de los mínimos cuadrados**, un método de ajuste de curvas que a principios del siglo XIX sugirió el matemático francés Adrien Legendre.

Para ilustrar esta técnica, consideremos los datos siguientes sobre el número de horas que 10 personas estudiaron para una prueba de francés y sus puntuaciones en la prueba:

Horas estudiadas x	*Puntuación en la prueba* y
4	31
9	58
10	65
14	73
4	37
7	44
12	60
22	91
1	21
17	84

Al hacer la gráfica de estos datos como en la figura 14.2, nos da la impresión de que una línea recta proporciona un ajuste razonablemente bueno. Aunque los puntos no caen todos en una línea recta, el patrón general sugiere que la puntuación promedio de la prueba para un número dado de horas de estudio bien puede estar relacionado con el número de horas estudiadas mediante una ecuación de la forma $\mu_{Y|x} = \alpha + \beta x$.

Una vez que hemos decidido en un problema dado que la regresión es aproximadamente lineal, nos enfrentamos al problema de estimar los coeficientes α y β de los datos muestrales. En otras palabras, nos enfrentamos al problema de obtener estimaciones de $\hat{\alpha}$ y $\hat{\beta}$ tales que la línea de regresión estimada $\hat{y} = \hat{\alpha} + \hat{\beta}x$ provea, en algún sentido, el mejor ajuste posible a los datos dados.

Al denotar la desviación vertical de un punto de la línea por e_i, como se indica en la figura 14.3, el criterio de los mínimos cuadrados sobre el cual basaremos esta "bondad de ajuste" requiere que minimicemos la suma de los cuadrados de estas desviaciones. Así, se nos da un conjunto de datos asociados en parejas $\{(x_i, y_i);\ i = 1, 2, \ldots, n\}$, las **estimaciones de mínimos cuadrados** de los coeficientes de regresión son los valores $\hat{\alpha}$ y $\hat{\beta}$ para los cuales la cantidad

$$q = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} [y_i - (\hat{\alpha} + \hat{\beta}x_i)]^2$$

es un mínimo. Al diferenciar parcialmente con respecto a $\hat{\alpha}$ y $\hat{\beta}$ y al igualar a cero estas derivadas parciales, obtenemos:

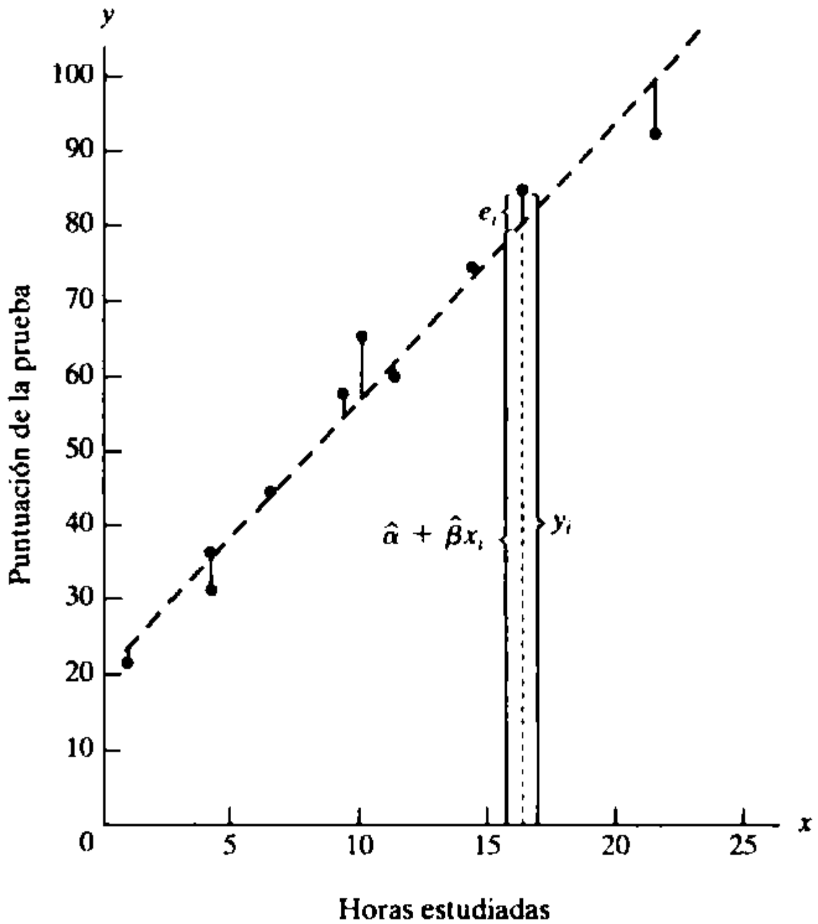

Figura 14.3 Criterio de mínimos cuadrados.

Entonces podemos escribir la estimación de mínimos cuadrados de α como

$$\hat{\alpha} = \frac{\sum_{i=1}^{n} y_i - \hat{\beta} \cdot \sum_{i=1}^{n} x_i}{n}$$

al resolver la primera de las dos ecuaciones normales para $\hat{\alpha}$. Esta fórmula para $\hat{\alpha}$ también se puede escribir como

$$\hat{\alpha} = \bar{y} - \hat{\beta} \cdot \bar{x}$$

Para simplificar la fórmula para $\hat{\beta}$ así como algunas de las fórmulas que encontraremos en las secciones 14.4 y 14.5, introduzcamos la notación siguiente:

$$S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2$$

$$S_{yy} = \sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2$$

y

$$S_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n} x_i y_i - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)$$

14.8 Dada la densidad conjunta

$$f(x, y) = \begin{cases} 24xy & \text{para } x > 0, y > 0 \text{ y } x + y < 1 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

demuestre que $\mu_{Y|x} = \frac{2}{3}(1 - x)$ y verifique este resultado al determinar los valores de $\mu_1, \mu_2, \sigma_1, \sigma_2$ y ρ y al sustituirlos en la primera fórmula del teorema 14.1.

14.9 Dada la densidad conjunta

$$f(x, y) = \begin{cases} 1 & \text{para } -y < x < y \text{ y } 0 < y < 1 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

demuestre que las variables aleatorias X y Y no están correlacionadas pero no son independientes.

14.10 Demuestre que si $\mu_{Y|x}$ es lineal en x y $\text{var}(Y|x)$ es constante, entonces $\text{var}(Y|x) = \sigma_2^2(1 - \rho^2)$.

14.11 Dado un par de variables aleatorias X y Y que tienen varianzas σ_1^2 y σ_2^2, y el coeficiente de correlación ρ, use el teorema 4.14 para expresar $\text{var}\left(\frac{X}{\sigma_1} + \frac{Y}{\sigma_2}\right)$ y $\text{var}\left(\frac{X}{\sigma_1} - \frac{Y}{\sigma_2}\right)$ en términos de σ_1, σ_2 y ρ. Después, al hacer uso del hecho que las varianzas no pueden ser negativas, demuestre que $-1 \leqq \rho \leqq +1$.

14.12 Dadas las variables aleatorias X_1, X_2 y X_3 que tienen la densidad conjunta $f(x_1, x_2, x_3)$, demuestre que si la regresión de X_3 sobre X_1 y X_2 es lineal y se escribe como

$$\mu_{X_3|x_1, x_2} = \alpha + \beta_1(x_1 - \mu_1) + \beta_2(x_2 - \mu_2)$$

entonces

$$\alpha = \mu_3$$

$$\beta_1 = \frac{\sigma_{13}\sigma_2^2 - \sigma_{12}\sigma_{23}}{\sigma_1^2\sigma_2^2 - \sigma_{12}^2}$$

$$\beta_2 = \frac{\sigma_{23}\sigma_1^2 - \sigma_{12}\sigma_{13}}{\sigma_1^2\sigma_2^2 - \sigma_{12}^2}$$

donde $\mu_i = E(X_i)$, $\sigma_i^2 = \text{var}(X_i)$ y $\sigma_{ij} = \text{cov}(X_i, X_j)$. [*Sugerencia*: proceda como en la página 454, al multiplicar por $(x_1 - \mu_1)$ y $(x_2 - \mu_2)$, respectivamente, para obtener la segunda y tercera ecuaciones.]

14.13 Encuentre la estimación de mínimos cuadrados del parámetro β en la ecuación de regresión $\mu_{Y|x} = \beta x$.

14.14 Resuelva simultáneamente las ecuaciones normales en la página 456 para demostrar que

$$\hat{\alpha} = \frac{\left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} y_i\right) - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} x_i y_i\right)}{n\left(\sum_{i=1}^{n} x_i^2\right) - \left(\sum_{i=1}^{n} x_i\right)^2}$$

(a) Encuentre la ecuación de la línea de mínimos cuadrados que nos permitirá predecir la puntuación del estudiante en el examen final en este curso sobre la base de su puntuación en el examen semestral.

(b) Prediga la puntuación del examen final de un estudiante que recibió 84 en el examen semestral.

14.18 La materia prima que se usa en la producción de una fibra sintética se almacena en un lugar que no tiene control de humedad. Las medidas de la humedad relativa y del contenido de humedad de muestras de la materia prima (ambas en porcentajes) en 12 días dieron los siguientes resultados:

Humedad	*Contenido de humedad*
46	12
53	14
37	11
42	13
34	10
29	8
60	17
44	12
41	10
48	15
33	9
40	13

(a) Ajuste una línea de mínimos cuadrados que nos permitirá predecir el contenido de humedad en términos de la humedad relativa.

(b) Use los resultados del inciso (a) para estimar (predecir) el contenido de humedad cuando la humedad relativa es del 38 por ciento.

14.19 Los siguientes datos corresponden al cloro residual en una alberca en diversos momentos después de haberse tratado con químicos:

Número de horas	*Cloro residual (partes por millón)*
2	1.8
4	1.5
6	1.4
8	1.1
10	1.1
12	0.9

$$\begin{aligned}\operatorname{var}(\hat{B}) &= \sum_{i=1}^{n}\left[\frac{x_i-\bar{x}}{S_{xx}}\right]^2 \cdot \operatorname{var}(Y_i|x_i)\\ &= \sum_{i=1}^{n}\left[\frac{x_i-\bar{x}}{S_{xx}}\right]^2 \cdot \sigma^2 = \frac{\sigma^2}{S_{xx}}\end{aligned}$$

Para aplicar esta teoría para probar hipótesis acerca de β o construir intervalos de confianza para β, tendremos que usar el siguiente teorema:

> **TEOREMA 14.3** Bajo las suposiciones del análisis de regresión normal, $\dfrac{n\hat{\sigma}^2}{\sigma^2}$ es un valor de una variable aleatoria que tiene la distribución ji cuadrada con $n-2$ grados de libertad. Además, esta variable aleatoria y $\hat{B}$ son independientes.

Al final de este capítulo damos una referencia de la demostración de este teorema.

Al hacer uso de este teorema así como del resultado probado anteriormente que $\hat{B}$ tiene una distribución normal con la media β y la varianza $\dfrac{\sigma^2}{S_{xx}}$, encontramos que la definición de la distribución t en la sección 8.5 nos lleva a

> **TEOREMA 14.4** Bajo las suposiciones del análisis de regresión normal,
>
> $$t = \frac{\dfrac{\hat{\beta}-\beta}{\sigma/\sqrt{S_{xx}}}}{\sqrt{\dfrac{n\hat{\sigma}^2}{\sigma^2}\Big/(n-2)}} = \frac{\hat{\beta}-\beta}{\hat{\sigma}}\sqrt{\frac{(n-2)S_{xx}}{n}}$$
>
> es un valor de una variable aleatoria que tiene la distribución t con $n-2$ grados de libertad.

Basado en esta estadística, probemos ahora una hipótesis acerca del coeficiente de regresión β.

EJEMPLO 14.5

Con respecto a los datos en la página 455 que corresponden a la cantidad de tiempo que 10 personas estudiaron para cierta prueba y a las puntuaciones que obtuvieron, prueba la hipótesis nula $\beta = 3$ contra la hipótesis alternativa $\beta > 3$ en el nivel 0.01 de significancia.

Solución

1. H_0: $\beta = 3$
   H_1: $\beta > 3$
   $\alpha = 0.01$

Solución

Al copiar las diversas cantidades de las páginas 458 y 467 y al sustituirlas junto con $t_{0.025, 8} = 2.306$ en la fórmula del intervalo de confianza del teorema 14.5, obtenemos

$$3.471 - (2.306)(4.720)\sqrt{\frac{10}{8(376)}} < \beta < 3.471 + (2.306)(4.720)\sqrt{\frac{10}{8(376)}}$$

o

$$2.84 < \beta < 4.10 \quad \blacktriangle$$

Puesto que los problemas de regresión de mayor complejidad en la realidad requieren cálculos bastante extensos, hoy en día se hacen prácticamente siempre con el software apropiado de computadoras. Una impresión así obtenida para nuestra ilustración se muestra en la figura 14.4; como se puede ver, proporciona no sólo los valores de $\hat{\alpha}$ y $\hat{\beta}$ en la columna encabezada COEFFICIENT, sino también estimaciones de las desviaciones estándar de las distribuciones muestrales de $\hat{A}$ y $\hat{B}$ en la columna encabezada ST. DEV. OF COEF. Si hubiésemos usado esta impresión en ejemplo 14.5, podríamos haber escrito el valor de la estadística t directamente como

$$t = \frac{3.471 - 3}{0.2723} = 1.73$$

y en el ejemplo 14.6 podríamos haber escrito los límites de confianza directamente como $3.471 \pm (2.306)(0.2723)$.

```
MTB  > NAME C1 = 'X'
MTB  > NAME C2 = 'Y'
MTB  > SET C1
DATA > 4   9   10   14   4   7   12   22   1   17
MTB  > SET C2
DATA > 31   58   65   73   37   44   60   91   21   84
MTB  > REGR C2 1 C1

  THE REGRESSION EQUATION IS
  Y = 21.7 + 3.47 X

                                     ST. DEV.     T-RATIO =
  COLUMN       COEFFICIENT           OF COEF.     COEF/S.D.
               21.693                  3.194          6.79
  X             3.4707                 0.2723        12.74
```

Figura 14.4 Impresión de computadora para ejemplos 14.4, 14.5 y 14.6.

EJERCICIOS

14.25 Haciendo uso del hecho que $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$ y $\hat{\beta} = \dfrac{S_{xy}}{S_{xx}}$, demuestre que

$$\sum_{i=1}^{n} [y_i - (\hat{\alpha} + \hat{\beta}x_i)]^2 = S_{yy} - \hat{\beta}S_{xy}$$

14.26 Demuestre que

(a) $\hat{\Sigma}^2$, la variable aleatoria que corresponde a $\hat{\sigma}^2$, no es un estimador insesgado de σ^2;

(b) $S_e^2 = \dfrac{n \cdot \hat{\Sigma}^2}{n-2}$ es un estimador insesgado de σ^2.

La cantidad s_e a menudo se conoce como el **error estándar de la estimación**.

14.27 Al usar s_e (véase ejercicio 14.26) en vez de $\hat{\sigma}$, reescriba

(a) la expresión para t en el teorema 14.4;

(b) la fórmula del intervalo de confianza del teorema 14.5.

14.28 Bajo las suposiciones del análisis de regresión normal, demuestre que

(a) la estimación de mínimos cuadrados de α se puede escribir en la forma

$$\hat{\alpha} = \sum_{i=1}^{n} \left[\frac{S_{xx} + n\bar{x}^2 - n\bar{x}x_i}{nS_{xx}}\right] y_i$$

(b) $\hat{\mathrm{A}}$ tiene una distribución normal con

$$E(\hat{\mathrm{A}}) = \alpha \quad \text{y} \quad \mathrm{var}(\hat{\mathrm{A}}) = \frac{(S_{xx} + n\bar{x}^2)\sigma^2}{nS_{xx}}$$

14.29 Use el teorema 14.15 para mostrar que

$$\mathrm{cov}(\hat{\mathrm{A}}, \hat{\mathrm{B}}) = -\frac{\bar{x}}{S_{xx}} \cdot \sigma^2$$

14.30 Use el resultado del inciso (b) del ejercicio 14.28 para demostrar que

$$z = \frac{(\hat{\alpha} - \alpha)\sqrt{nS_{xx}}}{\sigma\sqrt{S_{xx} + n\bar{x}^2}}$$

es un valor de una variable aleatoria que tiene la distribución normal estándar. También, use la primera parte del teorema 14.3 y el hecho que $\hat{\mathrm{A}}$ y $\dfrac{n\hat{\Sigma}^2}{\sigma^2}$ son independientes para demostrar que

$$t = \frac{(\hat{\alpha} - \alpha)\sqrt{(n-2)S_{xx}}}{\hat{\sigma}\sqrt{S_{xx} + n\bar{x}^2}}$$

es un valor de una variable aleatoria que tiene la distribución t con $n - 2$ grados de libertad.

14.31 Use los resultados del ejercicio 14.28 y 14.29 y el hecho que $E(\hat{B}) = \beta$ y $\text{var}(\hat{B}) = \dfrac{\sigma^2}{S_{xx}}$ para mostrar que $\hat{Y}_0 = \hat{A} + \hat{B}x_0$ es una variable aleatoria que tiene una distribución normal con la media

$$\alpha + \beta x_0 = \mu_{Y|x_0}$$

y la varianza

$$\sigma^2\left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right]$$

También, use la primera parte del teorema 14.3 así como el hecho que $\hat{Y}_0$ y $\dfrac{n\hat{\Sigma}^2}{\sigma^2}$ son independientes para demostrar que

$$t = \frac{(\hat{y}_0 - \mu_{Y|x_0})\sqrt{n-2}}{\hat{\sigma}\sqrt{1 + \dfrac{n(x_0 - \bar{x})^2}{S_{xx}}}}$$

es un valor de una variable aleatoria que tiene la distribución t con $n - 2$ grados de libertad.

14.32 Derive un intervalo con $(1 - \alpha)100\%$ de confianza para $\mu_{Y|x_0}$, la media de Y en $x = x_0$, al resolver la desigualdad doble $-t_{\alpha/2, n-2} < t < t_{\alpha/2, n-2}$ con t dada por la fórmula del ejercicio 14.31.

14.33 Use los resultados de los ejercicios 14.28 y 14.29 y el hecho que $E(\hat{B}) = \beta$ y $\text{var}(\hat{B}) = \dfrac{\sigma^2}{S_{xx}}$ para demostrar que $Y_0 - (\hat{A} + \hat{B}x_0)$ es una variable aleatoria que tiene la distribución normal con media cero y la varianza

$$\sigma^2\left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right]$$

En este caso Y_0 tiene una distribución normal con la media $\alpha + \beta x_0$ y la varianza σ^2; esto es, Y_0 es una observación futura de Y que corresponde a $x = x_0$. También, use la primera parte del teorema 14.3 así como el hecho que $Y_0 - (\hat{A} + \hat{B}x_0)$ y $\dfrac{n\hat{\Sigma}^2}{\sigma^2}$ son independientes para demostrar que.

$$t = \frac{[y_0 - (\hat{\alpha} + \hat{\beta}x_0)]\sqrt{n-2}}{\hat{\sigma}\sqrt{1 + n + \dfrac{n(x_0 - \bar{x})^2}{S_{xx}}}}$$

es un valor de una variable aleatoria que tiene la distribución t con $n - 2$ grados de libertad.

14.34 Resuelva la desigualdad doble $-t_{\alpha/2, n-2} < t < t_{\alpha/2, n-2}$ con t dada por la fórmula del ejercicio 14.33 de manera que el término medio es y_0 y los dos límites se

(a) Use el software apropiado de computadora para ajustar un línea recta a estos datos.

(b) Construya los límites con 99% de confianza para la pendiente de la línea ajustada.

14.47 Éstas son las cargas (gramos) que se colocaron en los extremos de varillas similares de plástico con las deflecciones resultantes (cm).

Carga x	*Deflección* y
25	1.58
30	1.39
35	1.41
40	1.60
55	1.81
45	1.78
50	1.65
60	1.94

(a) Use el software apropiado de computadora para ajustar una línea recta a estos datos.

(b) Use el nivel 0.95 de significancia para probar la hipótesis nula que $\beta = 0.01$ contra la alternativa que $\beta > 0.01$.

14.5 ANÁLISIS DE CORRELACIÓN NORMAL

En el análisis de correlación normal analizamos un conjunto de datos asociados en parejas $\{(x_i, y_i); i = 1, 2, \ldots, n\}$, donde las x_i las y_i son los valores de una muestra aleatoria de una población normal bivariada con los parámetros μ_1, μ_2, σ_1, σ_2 y ρ. Para estimar estos parámetros con el método de la máxima verosimilitud, tendremos que maximizar la verosimilitud

$$L = \prod_{i=1}^{n} f(x_i, y_i)$$

donde $f(x_i, y_i)$ esta dada por la definición 6.8, y con este fin tendremos que diferenciar L, o $\ln L$, parcialmente con respecto a μ_1, μ_2, σ_1, σ_2 y ρ, igualar las expresiones resultantes a cero, y después resolver el sistema resultante de ecuaciones para los cinco parámetros. Dejemos los detalles al lector y enunciemos meramente que cuando $\frac{\partial \ln L}{\partial \mu_1}$ y $\frac{\partial \ln L}{\partial \mu_2}$ se igualan a cero, obtenemos

$$-\frac{\sum_{i=1}^{n}(x_i - \mu_1)}{\sigma_1^2} + \frac{\rho \sum_{i=1}^{n}(y_i - \mu_2)}{\sigma_1 \sigma_2} = 0$$

y

$$-\frac{\rho \sum_{i=1}^{n}(x_i - \mu_1)}{\sigma_1 \sigma_2} + \frac{\sum_{i=1}^{n}(y_i - \mu_2)}{\sigma_2^2} = 0$$

$$S_{xx} = 771.35 - \frac{1}{10}(86.7)^2 = 19.661$$

$$S_{yy} = 819.34 - \frac{1}{10}(88.8)^2 = 30.796$$

$$S_{xy} = 792.92 - \frac{1}{10}(86.7)(88.8) = 23.024$$

y

$$r = \frac{23.024}{\sqrt{(19.661)(30.796)}} = 0.936$$

Esto es indicativo de una asociación positiva entre el tiempo que le toma a una secretaria ejecutar la tarea dada en la mañana y al final de la tarde, y esto es evidente en el **diagrama de dispersión** de la figura 14.5. Puesto que $100r^2 = 100(0.936)^2 = 87.6$, podemos decir que casi 88 por ciento de la variación de las y se explica mediante la relación lineal con x. ▲

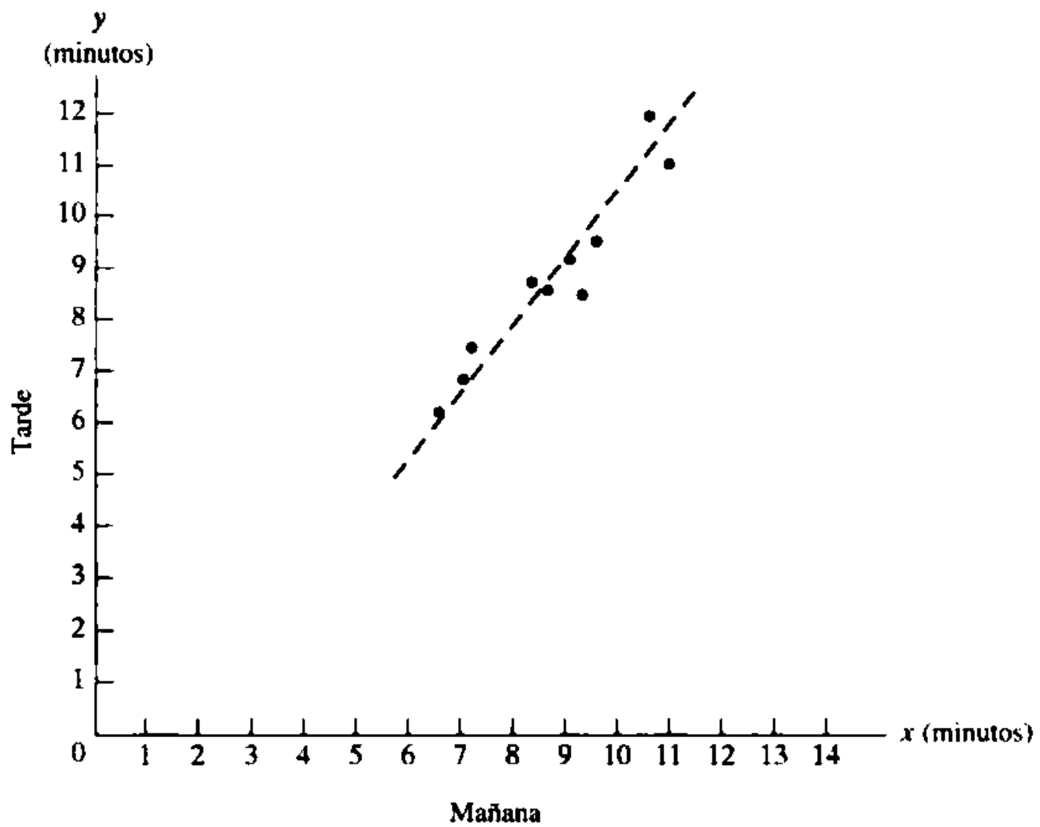

Figura 14.5 Diagrama de dispersión de los datos del ejemplo 14.7.

Puesto que la distribución muestral de R para muestras aleatorias de poblaciones normales bivariadas es más bien complicada, es práctica común basar los intervalos de confianza para ρ y las pruebas concernientes a ρ en la estadística:

$$\frac{1}{2} \cdot \ln \frac{1 + R}{1 - R}$$

cuya distribución es aproximadamente normal con la media $\frac{1}{2} \cdot \ln \frac{1 + \rho}{1 - \rho}$ y la varianza $\frac{1}{n - 3}$. Así,

$$z = \frac{\frac{1}{2} \cdot \ln \frac{1 + r}{1 - r} - \frac{1}{2} \cdot \ln \frac{1 + \rho}{1 - \rho}}{\frac{1}{\sqrt{n - 3}}}$$

$$= \frac{\sqrt{n - 3}}{2} \cdot \ln \frac{(1 + r)(1 - \rho)}{(1 - r)(1 + \rho)}$$

se puede considerar como un valor de una variable aleatoria que tiene aproximadamente la distribución normal estándar. Al usar esta aproximación, podemos probar la hipótesis nula $\rho = \rho_0$ contra una alternativa apropiada, como se ilustra en el ejemplo 14.8, o calcular los intervalos de confianza para ρ mediante el método sugerido en el ejercicio 14.51.

EJEMPLO 14.8

Con respecto al ejemplo 14.7, pruebe la hipótesis nula $\rho = 0$ contra la hipótesis alternativa $\rho \neq 0$ en el nivel 0.01 de significancia.

Solución

1. H_0: $\rho = 0$
   H_1: $\rho \neq 0$
   $\alpha = 0.01$
2. Rechace la hipótesis nula si $z \leqq -2.575$ o $z \geqq 2.575$, donde

$$z = \frac{\sqrt{n - 3}}{2} \cdot \ln \frac{1 + r}{1 - r}$$

3. Al sustituir $n = 10$ y $r = 0.936$, obtenemos

$$z = \frac{\sqrt{7}}{2} \cdot \ln \frac{1.936}{0.064} = 4.5$$

4. Puesto que $z = 4.5$ excede a 2.575, debemos rechazar la hipótesis nula; concluimos que hay una relación lineal entre el tiempo que tarda una secretaria en llenar el formulario en la mañana y al final de la tarde. ▲

EJERCICIOS

14.48 Verifique que la fórmula para t del teorema 14.4 se puede escribir como

$$t = \left(1 - \frac{\beta}{\hat{\beta}}\right)\frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

14.49 Use la fórmula para t del ejercicio 14.48 para derivar los límites con $(1-\alpha)100\%$ de confianza siguientes para β:

$$\hat{\beta}\left[1 \pm t_{\alpha/2,n-2}\cdot\frac{\sqrt{1-r^2}}{r\sqrt{n-2}}\right]$$

14.50 Use la fórmula para t del ejercicio 14.48 para demostrar que si las suposiciones que sustentan el análisis de regresión normal se satisfacen y $\beta = 0$, entonces R^2 tiene la distribución beta con la media $\dfrac{1}{n-1}$.

14.51 Al resolver la desigualdad doble $-z_{\alpha/2} \leqq z \leqq z_{\alpha/2}$ (con z dada por la fórmula en la página 477) para ρ, derive una fórmula para un intervalo con $(1-\alpha)100\%$ de confianza para ρ.

14.52 En una muestra aleatoria de n pares de valores X y Y, (x_i, y_j) ocurre f_{ij} veces para $i = 1, 2, \ldots, r$ y $j = 1, 2, \ldots, c$. Sea que $f_{i\cdot}$ denote el número de pares donde X asume el valor x_i y $f_{\cdot j}$ el número de pares donde Y asume el valor y_j, escriba una fórmula para el coeficiente de correlación.

APLICACIONES

14.53 Se dice que una prueba de rendimiento es confiable si un estudiante que tome la prueba varias veces obtendrá consistentemente puntuaciones altas (o bajas). Una forma de verificar la confiabilidad de una prueba es dividirla en dos partes, por lo general los problemas con numeración par y los problemas con numeración impar, y observar la correlación entre las puntuaciones que los estudiantes obtienen en ambas mitades de la prueba. Así, los datos siguientes representan las calificaciones, x y y, que 20 estudiantes obtuvieron para los problemas con numeración par y los problemas con numeración impar de una nueva prueba objetiva diseñada para probar el rendimiento de alumnos del último año de primaria en ciencias en general:

x	y	x	y
27	29	33	42
36	44	39	31
44	49	38	38
32	27	24	22
27	35	33	34
41	33	32	37
38	29	37	38
44	40	33	35
30	27	34	32
27	38	39	43

Calcule r para estos datos y pruebe su significancia, esto es, la hipótesis nula $\rho = 0$ contra la hipótesis alternativa $\rho \neq 0$ en el nivel 0.05 de significancia.

14.54 Con respecto al ejercicio 14.53, use la fórmula obtenida en el ejercicio 14.51 para construir un intervalo con 95 por ciento de confianza para ρ.

14.55 Los datos siguientes corresponden a x, la cantidad de fertilizante (en libras) que un agricultor aplica a su suelo, y y, es su rendimiento de trigo (en "bushels" por acre):

x	y	x	y	x	y
112	33	88	24	37	27
92	28	44	17	23	9
72	38	132	36	77	32
66	17	23	14	142	38
112	35	57	25	37	13
88	31	111	40	127	23
42	8	69	29	88	31
126	37	19	12	48	37
72	32	103	27	61	25
52	20	141	40	71	14
28	17	77	26	113	26

Suponga que los datos se pueden considerar como una muestra aleatoria de una población bivariada normal, calcule r y pruebe su significancia en el nivel 0.01 de significancia. También, dibuje un diagrama de dispersión de estos datos asociados en parejas y juzgue si la suposición parece razonable.

14.56 Con respecto al ejercicio 14.55, use la fórmula obtenida en el ejercicio 14.51 para construir un intervalo de confianza del 99% para ρ.

14.57 Use la fórmula del ejercicio 14.48 para calcular un intervalo de confianza del 95% para β para los números de horas de estudio y las puntuaciones de la prueba en la página 455, y compare este intervalo con el obtenido en el ejemplo 14.6.

14.58 A menudo se puede simplificar el cálculo de r al sumar la misma constante a cada x, sumar la misma constante a cada y, o multiplicar cada x y/o y por las mismas constantes positivas. Vuelva a calcular r para los datos del ejemplo 14.7, multiplique primero cada x y cada y por 10 y después reste 70 de cada x y 60 de cada y.

14.59 La tabla en la página 480 muestra cómo se distribuyen las puntuaciones en historia y economía de 25 estudiantes. Use el método del ejercicio 14.52 para determinar el valor de r, reemplace los encabezados de los renglones con las **marcas de clase** correspondientes (puntos medios) 23, 28, 33, 38, 43 y 48 y los encabezados de las columnas con las marcas de clase correspondientes 23, 28, 33, 38 y 43. Use este valor de r para probar en el nivel 0.05 de significancia si hay una relación entre las puntuaciones en las dos materias.

14.60 Rehaga el ejercicio 14.59, codifique las marcas de clase de las puntuaciones de historia $-2, -1, 0, 1$ y 2 y las marcas de clase de las puntuaciones de economía

Puntuaciones en economía \ *Puntuaciones en historia*	21–25	26–30	31–35	36–40	41–45
21–25	1				
26–30		3	1		
31–35		2	5	2	
36–40			1	4	1
41–45			1	3	
46–50					1

−2, −1, 0, 1, 2 y 3. (Se sigue, por el ejercicio 14.58, que esta clase de codificación no afectará el valor de r.)

14.61 Si las categorías de renglón así como las categorías de columna de una tabla $r \times c$ están ordenadas, podemos reemplazar los encabezados de los renglones y también los encabezados de las columnas con enteros consecutivos y después volver a calcular r con la fórmula obtenida en el ejercicio 14.52. Use este método para rehacer el ejemplo 13.11, reemplace Bajo, Promedio y Alto en cada caso por −1, 0 y 1.

14.62 Con respecto a la segunda tabla $r \times c$ en la página 438, use el método sugerido en el ejercicio 14.61 para probar en el nivel 0.05 de significancia si hay una relación entre el IQ y el desempeño en el trabajo. Reemplace los encabezados de los renglones así como los encabezados de las columnas con −1, 0 y 1.

14.63 (a) Use un programa apropiado de computadora para obtener el coeficiente de correlación muestral para los datos del ejercicio 14.46.
(b) Pruebe si r es significativamente diferente de 0, use el nivel 0.05.

14.64 (a) Use un programa apropiado de computadora para obtener el coeficiente de correlación muestral para los datos del ejercicio 14.47.
(b) Pruebe si este coeficiente es significante, use el nivel 0.10.

14.6 REGRESIÓN LINEAL MÚLTIPLE

Aunque hay muchos problemas donde una variable se puede predecir con bastante exactitud en términos de otras, es lógico que los pronósticos se deben mejorar si uno considera información pertinente adicional. Por ejemplo, debemos poder hacer mejores predicciones del desempeño de profesores recién contratados si consideramos no sólo su educación, sino también los años de experiencia y su personalidad. También, debemos poder hacer mejores pronósticos del éxito de un nuevo libro de texto si consideramos no sólo la calidad del trabajo, sino también la demanda potencial y la competencia.

Aunque se pueden usar muchas fórmulas diferentes para expresar las relaciones de regresión entre más de dos variables (véase, por ejemplo, el ejemplo 14.3), las más ampliamente usadas son las ecuaciones lineales de la forma:

14.7 REGRESIÓN LINEAL MÚLTIPLE (NOTACIÓN MATRICIAL)†

El modelo que estamos usando en la regresión lineal múltiple se presta de manera única a un tratamiento unificado en notación matricial. Esta notación hace posible enunciar resultados generales en forma compacta y utilizar muchos resultados de la teoría matricial con gran ventaja. Como es costumbre, denotaremos las matrices con letras mayúsculas en tipo negritas.

Podríamos introducir el enfoque matricial al expresar la suma de los cuadrados q (que minimizamos en la sección anterior al diferenciar parcialmente con respecto a las $\hat{\beta}$) en notación matricial y arrancar de ahí, pero dejamos esto al lector en el ejercicio 14.65; empecemos aquí con las ecuaciones normales en la página 482.

Para expresar las ecuaciones normales en notación matricial, definamos las tres matrices siguientes:

$$\mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix}$$

$$\mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \quad \text{y} \quad \mathbf{B} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{pmatrix}$$

La primera $\mathbf{X}$ es una matriz de $n \times (k + 1)$ que consiste esencialmente de los valores dados de las x, donde se añade una columna 1 para dar cabida a los términos constantes. $\mathbf{Y}$ es una matriz de $n \times 1$ (o vector columna) que consiste en los valores observados de Y, y $\mathbf{B}$ es una matriz $(k + 1) \times 1$ (o vector columna) que consiste en las estimaciones de mínimos cuadrados de los coeficientes de regresión.

Al usar estas matrices, podemos ahora escribir la siguiente solución simbólica de las ecuaciones normales en la página 482.

> **TEOREMA 14.7** Las estimaciones de mínimos cuadrados para los coeficientes de regresión múltiple están dadas por
>
> $$\mathbf{B} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$
>
> donde $\mathbf{X}'$ es la transpuesta de $\mathbf{X}$ y $(\mathbf{X}'\mathbf{X})^{-1}$ es la inversa de $\mathbf{X}'\mathbf{X}$.

† Para esta sección se supone que el lector está familiarizado con el material normalmente cubierto en un primer curso de álgebra matricial. Puesto que la notación matricial no se usa en ninguna otra parte de este libro, esta sección se puede omitir sin pérdida de continuidad.

Demostración. Primero determinamos $\mathbf{X}'\mathbf{X}$, $\mathbf{X}'\mathbf{X}\mathbf{B}$, y $\mathbf{X}'\mathbf{Y}$, y obtenemos

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} n & \sum x_1 & \sum x_2 & \cdots & \sum x_k \\ \sum x_1 & \sum x_1^2 & \sum x_1 x_2 & \cdots & \sum x_1 x_k \\ \sum x_2 & \sum x_2 x_1 & \sum x_2^2 & \cdots & \sum x_2 x_k \\ & & \cdots & & \\ \sum x_k & \sum x_k x_1 & \sum x_k x_2 & \cdots & \sum x_k^2 \end{pmatrix}$$

$$\mathbf{X}'\mathbf{X}\mathbf{B} = \begin{pmatrix} \hat{\beta}_0 \cdot n + \hat{\beta}_1 \cdot \sum x_1 + \hat{\beta}_2 \cdot \sum x_2 + \cdots + \hat{\beta}_k \cdot \sum x_k \\ \hat{\beta}_0 \cdot \sum x_1 + \hat{\beta}_1 \cdot \sum x_1^2 + \hat{\beta}_2 \cdot \sum x_1 x_2 + \cdots + \hat{\beta}_k \cdot \sum x_1 x_k \\ \hat{\beta}_0 \cdot \sum x_2 + \hat{\beta}_1 \cdot \sum x_2 x_1 + \hat{\beta}_2 \cdot \sum x_2^2 + \cdots + \hat{\beta}_k \cdot \sum x_2 x_k \\ \cdots \\ \hat{\beta}_0 \cdot \sum x_k + \hat{\beta}_1 \cdot \sum x_k x_1 + \hat{\beta}_2 \cdot \sum x_k x_2 + \cdots + \hat{\beta}_k \cdot \sum x_k^2 \end{pmatrix}$$

$$\mathbf{X}'\mathbf{Y} = \begin{pmatrix} \sum y \\ \sum x_1 y \\ \sum x_2 y \\ \cdots \\ \sum x_k y \end{pmatrix}$$

Al identificar los elementos de $\mathbf{X}'\mathbf{X}\mathbf{B}$ como las expresiones en el lado derecho de las ecuaciones normales en la página 482 y las de $\mathbf{X}'\mathbf{Y}$ como las expresiones en el lado izquierdo, podemos escribir

$$\mathbf{X}'\mathbf{X}\mathbf{B} = \mathbf{X}'\mathbf{Y}$$

Al multiplicar en el lado izquierdo por $(\mathbf{X}'\mathbf{X})^{-1}$, obtenemos

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\mathbf{B} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

y finalmente

$$\mathbf{B} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

puesto que $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}$ es igual a la matriz de identidad $\mathbf{I}$ $(k+1) \times (k+1)$ y por definición $\mathbf{IB} = \mathbf{B}$. En este caso hemos supuesto que $\mathbf{X}'\mathbf{X}$ no tiene singularidad de manera que existe su inversa. ▼

EJEMPLO 14.11

Con respecto al ejemplo 14.9, use el teorema 14.7 para determinar las estimaciones de mínimos cuadrados de los coeficientes de regresión múltiple.

Solución

Al sustituir $\sum x_1 = 25$, $\sum x_2 = 16$, $\sum x_1^2 = 87$, $\sum x_1 x_2 = 55$, $\sum x_2^2 = 36$ y $n = 8$ de la página 482 en la expresión para $\mathbf{X}'\mathbf{X}$ de arriba, obtenemos

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} 8 & 25 & 16 \\ 25 & 87 & 55 \\ 16 & 55 & 36 \end{pmatrix}$$

Entonces, la inversa de esta matriz se puede obtener mediante cualquiera de diversas técnicas; al usar la que está basada en los cofactores, encontramos que

$$(\mathbf{X}'\mathbf{X})^{-1} = \frac{1}{84}\cdot\begin{pmatrix} 107 & -20 & -17 \\ -20 & 32 & -40 \\ -17 & -40 & 71 \end{pmatrix}$$

donde 84 es el valor de $|\mathbf{X}'\mathbf{X}|$, el determinante de $\mathbf{X}'\mathbf{X}$.

Al sustituir $\sum y = 637{,}000$, $\sum x_1 y = 2{,}031{,}100$ y $\sum x_2 y = 1{,}297{,}700$ de la página 482 en la expresión para $\mathbf{X}'\mathbf{Y}$ en la página 485, obtenemos entonces

$$\mathbf{X}'\mathbf{Y} = \begin{pmatrix} 637{,}000 \\ 2{,}031{,}100 \\ 1{,}297{,}700 \end{pmatrix}$$

y finalmente

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \frac{1}{84}\cdot\begin{pmatrix} 107 & -20 & -17 \\ -20 & 32 & -40 \\ -17 & -40 & 71 \end{pmatrix}\begin{pmatrix} 637{,}000 \\ 2{,}031{,}100 \\ 1{,}297{,}700 \end{pmatrix}$$

$$= \frac{1}{84}\cdot\begin{pmatrix} 5{,}476{,}100 \\ 347{,}200 \\ 63{,}700 \end{pmatrix}$$

$$= \begin{pmatrix} 65{,}191.7 \\ 4{,}133.3 \\ 758.3 \end{pmatrix}$$

donde las $\hat{\beta}$ están redondeadas a un decimal. Advierta que los resultados obtenidos aquí son idénticos a los mostrados en la impresión de computadora de la figura 14.6. ▲

A continuación, para generalizar el trabajo de la sección 14.4, hacemos suposiciones que son muy similares a las de la página 464: suponemos que para $i = 1, 2, \ldots$ y n, las Y_i son variables aleatorias independientes que tienen distribuciones normales con las medias $\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}$ y la desviación estándar común σ. Con base en n puntos de datos

$$(x_{i1}, x_{i2}, \ldots, x_{ik}, y_i)$$

podemos entonces hacer toda clase de inferencias sobre los parámetros de nuestro modelo, las β y σ, y juzgar los méritos de las estimaciones y las predicciones basadas en la ecuación estimada de regresión múltiple.

Encontrar las estimaciones de máxima verosimilitud de las β y σ es directo, como en las páginas 464 y 465, y se dejará al lector en el ejercicio 14.65. Los resultados son como sigue: las estimaciones de máxima verosimilitud de las β son iguales a las estimaciones correspondientes de mínimos cuadrados, así que están dadas por los elementos de la matriz columna $(k + 1) \times 1$.

$$\mathbf{B} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

La estimación de máxima verosimilitud de σ está dada por

$$\hat{\sigma} = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} [y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \cdots + \hat{\beta}_k x_{ik})]^2}$$

donde las $\hat{\beta}$ son las estimaciones de máxima verosimilitud de las β y, como se pedirá al lector que verifique en el ejercicio 14.67, también se puede escribir como

$$\hat{\sigma} = \sqrt{\frac{\mathbf{Y}'\mathbf{Y} - \mathbf{B}'\mathbf{X}'\mathbf{Y}}{n}}$$

en notación matricial.

EJEMPLO 14.12

Use los resultados del ejemplo 14.11 para determinar el valor de $\hat{\sigma}$ para los datos del ejemplo 14.9.

Solución

Calculemos primero $\mathbf{Y}'\mathbf{Y}$, lo cual es simplemente $\sum_{i=1}^{n} y_i^2$, así obtenemos

$$\begin{aligned}\mathbf{Y}'\mathbf{Y} &= 78{,}800^2 + 74{,}300^2 + \cdots + 82{,}900^2 \\ &= 50{,}907{,}080{,}000\end{aligned}$$

Entonces, al copiar $\mathbf{B}$ y $\mathbf{X}'\mathbf{Y}$ de la página 486, obtenemos

$$\begin{aligned}\mathbf{B}'\mathbf{X}'\mathbf{Y} &= \frac{1}{84} \cdot (5{,}476{,}100 \quad 347{,}200 \quad 63{,}700)\begin{pmatrix} 637{,}000 \\ 2{,}031{,}100 \\ 1{,}297{,}700 \end{pmatrix} \\ &= 50{,}906{,}394{,}166\end{aligned}$$

y se sigue que

$$\begin{aligned}\hat{\sigma} &= \sqrt{\frac{50{,}907{,}080{,}000 - 50{,}906{,}394{,}166}{8}} \\ &= 292.8 \quad \blacktriangle\end{aligned}$$

Es interesante observar que la estimación que hemos obtenido aquí no es igual a la que se muestra en la impresión de computadora de la figura 14.6. La estimación que ahí se muestra, $S = 370.4$, es tal que S^2 es un estimador insesgado de σ^2, análogo al error estándar de la estimación que definimos en la página 469. Difiere de $\hat{\sigma}$ en que dividimos por $n - k - 1$ en vez de n, y si hubiésemos hecho esto en nuestro ejemplo, habríamos obtenido

$$\begin{aligned}s_e &= \sqrt{\frac{50{,}907{,}080{,}000 - 50{,}906{,}394{,}166}{8 - 2 - 1}} \\ &= 370.4\end{aligned}$$

14.71 Si $x_{01}, x_{02}, \ldots, x_{0k}$ son valores dados de $x_1, x_2, \ldots, x_k$ y $\mathbf{X}_0$ es el vector columna

$$\mathbf{X}_0 = \begin{pmatrix} 1 \\ x_{01} \\ x_{02} \\ \cdots \\ x_{0k} \end{pmatrix}$$

se puede mostrar que

$$t = \frac{\mathbf{B}'\mathbf{X}_0 - \mu_{Y|x_{01}, x_{02}, \ldots, x_{0k}}}{\hat{\sigma} \cdot \sqrt{\dfrac{n[\mathbf{X}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_0]}{n-k-1}}}$$

es un valor de una variable aleatoria que tiene la distribución t con $n - k - 1$ grados de libertad.

(a) Muestre que para $k = 1$ esta estadística es equivalente a la del ejercicio 14.31.

(b) Derive una fórmula para un intervalo $(1 - \alpha)100\%$ de confianza para

$$\mu_{Y|x_{01}, x_{02}, \ldots, x_{0k}}$$

14.72 Con $x_{01}, x_{02}, \ldots, x_{0k}$ y $\mathbf{X}_0$ como se definieron en el ejercicio 14.71 y Y_0 es una variable aleatoria que tiene una distribución normal con la media $\beta_0 + \beta_1 x_{01} + \cdots + \beta_k x_{0k}$ y la varianza σ^2, se puede demostrar que

$$t = \frac{y_0 - \mathbf{B}'\mathbf{X}_0}{\hat{\sigma} \cdot \sqrt{\dfrac{n[1 + \mathbf{X}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_0]}{n-k-1}}}$$

es un valor de una variable aleatoria que tiene la distribución t con $n - k - 1$ grados de libertad.

(a) Demuestre que para $k = 1$ esta estadística es equivalente a la del ejercicio 14.33.

(b) Derive una fórmula para límites de predicción de $(1 - \alpha)100\%$ para una observación futura de Y_0.

APLICACIONES

14.73 Éstos son datos muestrales proporcionados por una compañía de mudanzas sobre los pesos de seis embarques, las distancias que se trasladaron, y el daño en que se incurrió:

Peso (1,000 *libras*) x_1	*Distancia* (1,000 *millas*) x_2	*Daño* (*dólares*) y
4.0	1.5	160
3.0	2.2	112
1.6	1.0	69
1.2	2.0	90
3.4	0.8	123
4.8	1.6	186

Dureza (*Rockwell 30-T*) y	*Contenido de cobre* (*por ciento*) x_1	*Temperatura de templado* (*grados F*) x_2
78.9	0.02	1,000
55.2	0.02	1,200
80.9	0.10	1,000
57.4	0.10	1,200
85.3	0.18	1,000
60.7	0.18	1,200

Ajuste un plano con el método de los mínimos cuadrados y úselo para estimar la dureza promedio de esta clase de acero cuando el contenido de cobre es 0.14 por ciento y la temperatura de templado es 1,100°F.

14.77 Cuando las $x_1, x_2, \ldots,$ y/o las x_k están uniformemente espaciadas, el cálculo de las $\hat{\beta}$ se puede simplificar al usar la codificación sugerida en el ejercicio 14.15. Vuelva a resolver el ejercicio 14.76, codifique los valores de x_1 como -1, 0, y 1 y los valores de x_2 como -1 y 1. (Observe que para las x_1 y x_2, codificadas, llámelas z_1 y z_2, tenemos no sólo $\sum z_1 = 0$ y $\sum z_2 = 0$, sino también $\sum z_1 z_2 = 0$.)

14.78 Los siguientes son datos sobre el porcentaje de efectividad de un analgésico y las cantidades de tres diferentes medicamentos (en miligramos) presentes en cada cápsula:

Medicamento A x_1	*Medicamento B* x_2	*Medicamento C* x_3	*Porcentaje de eficacia* y
15	20	10	47
15	20	20	54
15	30	10	58
15	30	20	66
30	20	10	59
30	20	20	67
30	30	10	71
30	30	20	83
45	20	10	72
45	20	20	82
45	30	10	85
45	30	20	94

Suponga que la regresión es lineal, estime los coeficientes de regresión después de codificar apropiadamente cada una de las x, y exprese la ecuación de regresión estimada en términos de las variables originales.

14.79 Los modelos de regresión que introdujimos en las secciones 14.2 y 14.6 son lineales en las x, pero, lo que es más importante, también son lineales en las β.

Ciertamente, se pueden usar en algunos problemas donde la relación entre las x y la y no es lineal. Por ejemplo, cuando la regresión es parabólica y de la forma

$$\mu_{Y|x} = \beta_0 + \beta_1 x + \beta_2 x^2$$

simplemente usamos la ecuación de regresión $\mu_{Y|x} = \beta_0 + \beta_1 x_1 + \beta_2 x_2$ con $x_1 = x$ y $x_2 = x^2$. Use este método para ajustar una parábola a los datos siguientes sobre el tiempo de secado de un barniz y la cantidad de cierto producto químico que se le ha añadido:

Cantidad de aditivo (gramos) x	*Tiempo de secado (horas)* y
1	8.5
2	8.0
3	6.0
4	5.0
5	6.0
6	5.5
7	6.5
8	7.0

También, prediga el tiempo de secado cuando se añaden 6.5 gramos del producto químico.

14.80 Los siguientes datos corresponden a la demanda de un producto (en miles de unidades) y el precio (en centavos) que se cobró en cinco diferentes áreas de mercado:

Precio x	*Demanda* y
20	22
16	41
10	120
11	89
14	56

Ajuste una parábola a estos datos por el método sugerido en el ejercicio 14.79.

14.81 Para juzgar si vale la pena ajustar una parábola en el ejercicio 14.80 y no sólo una línea recta, pruebe la hipótesis nula $\beta_2 = 0$ contra la hipótesis alternativa $\beta_2 \neq 0$ en el nivel 0.05 de significancia.

14.82 Use los resultados obtenidos para los datos del ejemplo 14.9 a fin de construir un intervalo de confianza del 90% para el coeficiente de regresión β_2 (véase el ejercicio 14.70).

14.83 Con respecto al ejercicio 14.73, pruebe la hipótesis nula $\beta_2 = 10.0$ contra la hipótesis alternativa $\beta_2 \neq 10.0$ en el nivel 0.05 de significancia.

14.84 Con respecto al ejercicio 14.73, construya un intervalo de confianza del 95% para el coeficiente de regresión β_1.

14.85 Con respecto al ejercicio 14.74, pruebe la hipótesis nula $\beta_1 = 0.12$ contra la hipótesis alternativa $\beta_1 < 0.12$ en el nivel 0.05 de significancia.

14.86 Con respecto al ejercicio 14.74, construya un intervalo de confianza del 98% para el coeficiente de regresión β_2.

14.87 Use los resultados obtenidos para los datos del ejemplo 14.9 y el resultado del inciso (b) del ejercicio 14.71 para construir un intervalo de confianza del 95% para la media del precio de venta de una casa de tres recámaras con dos baños en el desarrollo habitacional dado.

14.88 Use los resultados obtenidos para los datos del ejemplo 14.9 y el resultado del inciso (b) del ejercicio 14.72 para construir límites de predicción del 99% para el precio de venta de una casa de tres recámaras con dos baños en el desarrollo habitacional dado.

14.89 Con respecto al ejercicio 14.73, use el resultado del inciso (b) del ejercicio 14.71 para construir un intervalo de confianza del 98% para la media del daño de embarques de 2,400 libras que se trasladan 1,200 millas.

14.90 Con respecto al ejercicio 14.73, use el resultado del inciso (b) del ejercicio 14.72 para construir límites de pronóstico del 95% para el daño en que se incurrirá en un embarque de 2,400 libras que se traslada 1,200 millas.

14.91 Con respecto al ejercicio 14.74, use el resultado del inciso (b) del ejercicio 14.71 para construir un intervalo de confianza del 99% para el promedio de la utilidad neta semanal de restaurantes con 210 lugares en una localidad donde la cuenta del tráfico diario promedia 14,000 autos.

14.92 Con respecto al ejercicio 14.74, use el resultado del inciso (b) del ejercicio 14.72 para construir límites de predicción del 98% para el promedio de la utilidad neta semanal de restaurantes con 210 lugares en una localidad donde la cuenta del tráfico diario promedia 14,000 autos.

14.93 Use un programa apropiado de computadora para rehacer el ejercicio 14.78 sin codificar los valores x.

14.94 (a) Use un programa apropiado de computadora para ajustar un plano a los datos siguientes relativos al uso mensual de agua de una planta de producción (galones) a su producción mensual (toneladas), la media de la temperatura ambiente mensual (°F), y el número mensual de días de operación de la planta durante un periodo de 12 meses.

Uso de agua y	*Produción* x_1	*Media de la temperatura* x_2	*Días de operación* x_3
2,228	98.5	67.4	19
2,609	108.2	70.3	20
3,088	109.6	82.1	21
2,378	101.0	69.2	21
1,980	83.3	64.5	19
1,717	70.0	63.7	21
2,723	144.7	58.0	19
2,031	84.4	58.1	20
1,902	97.4	36.6	17
1,721	131.8	49.6	23
2,254	82.1	44.3	18
2,522	64.5	44.1	19

(b) Estime el uso de agua de la planta durante un mes cuando su producción es 90.0 toneladas, la media de la temperatura ambiente es 65°F, y opera por 20 días.

REFERENCIAS

En el libro de S. S. Wilks, al que se hizo referencia al final del capítulo 7, se puede encontrar una demostración del teorema 14.3 y otros detalles matemáticos que se omitieron en el texto, y en el libro de Kendall y Stuart, al que se hizo referencia al final del capítulo 3, se encuentra información sobre la distribución de $\frac{1}{2} \cdot \ln \frac{1+R}{1-R}$. Una derivación de las estimaciones de máxima verosimilitud de σ_1, σ_2 y ρ se da en la tercera edición (pero no en la cuarta) de

HOEL, P., *Introduction to Mathematical Statistics*, 3a. ed. Nueva York: John Wiley & Sons, Inc., 1962.

Tratamientos más detallados de la regresión múltiple se pueden encontrar en muchos libros más avanzados, por ejemplo, en

MORRISON, D. F., *Applied Linear Statistical Methods*. Upper Saddle River, N.J.: Prentice Hall, 1983,

WEISBERG, S., *Applied Linear Regression*, 2a. ed. Nueva York: John Wiley & Sons, Inc., 1985,

WONNACOTT, T. H., and WONNACOTT, R. J., *Regression: A Second Course in Statistics*. Nueva York: John Wiley & Sons, Inc., 1981.

CAPÍTULO

15

Análisis de la varianza

15.1 INTRODUCCIÓN

En este capítulo generalizaremos el trabajo de la sección 13.3 y consideraremos el problema de decidir si las diferencias observadas entre más de dos medias muestrales se puede atribuir al azar o si hay diferencias reales entre las medias de las poblaciones muestreadas. Por ejemplo, quizá deseamos decidir con base en datos muestrales si realmente hay una diferencia en la eficacia de tres métodos de enseñar una lengua extranjera, tal vez queremos comparar los rendimientos promedio por acre de seis variedades de trigo, o deseamos ver si realmente hay diferencia en el millaje promedio obtenido con cuatro clases de gasolina.

Puesto que las diferencias observadas siempre se pueden deber a causas distintas a las postuladas; por ejemplo, las diferencias en el desempeño de los estudiantes a quienes se les enseña una lengua extranjera mediante tres métodos diferentes se pueden deber a diferencias en inteligencia, y las diferencias en el millaje promedio obtenido con cuatro clases de gasolina se puede deber a las diferencias en las condiciones del camino; también examinaremos algunos puntos del **diseño de experimentos** de manera que, con seguridad razonable, los resultados estadísticamente significativos se puedan atribuir a causas específicas.

15.2 ANÁLISIS DE LA VARIANZA EN UN SENTIDO

Para dar un ejemplo de una situación típica donde haríamos un análisis de la varianza en un sentido, suponga que queremos comparar la acción limpiadora de tres detergentes con base en las siguientes lecturas de blancura en 15 muestras de tela blanca, que primero se mancharon con tinta china y después se lavaron en una máquina tipo agitador con los detergentes respectivos

Detergente A :	77, 81, 71, 76, 80
Detergente B :	72, 58, 74, 66, 79
Detergente C :	76, 85, 82, 80, 77

Las medias de estas tres muestras fueron 77, 68 y 80, y queremos saber si las diferencias entre ellas son significativas o si se pueden atribuir al azar.

En general, en un problema como éste, tenemos muestras aleatorias independientes de tamaño n de k poblaciones. El $j$ésimo valor de la $i$ésima población se denota conx_{ij}, esto es,

Población 1:	$x_{11}, x_{12}, \ldots, x_{1n}$
Población 2:	$x_{21}, x_{22}, \ldots, x_{2n}$
. . .	
Población k:	$x_{k1}, x_{k2}, \ldots, x_{kn}$

Y supondremos que las variables aleatorias correspondientes X_{ij}, que son todas independientes, tienen distribuciones normales con las respectivas medias μ_i y la varianza común σ^2. Al enunciar estas suposiciones de una manera algo diferente, podríamos decir que el modelo para las observaciones está dado por

$$x_{ij} = \mu_i + e_{ij}$$

para $i = 1, 2, \ldots, k$ y $j = 1, 2, \ldots, n$, donde e_{ij} son los valores de nk variables aleatorias independientes que tienen distribuciones normales con medias cero y la varianza común σ^2. Para permitir la generalización de este modelo a clases de situaciones más complicadas (véanse las páginas 506 y 507), suelen escribirse en la forma

$$x_{ij} = \mu + \alpha_i + e_{ij}$$

para $i = 1, 2, \ldots, k$ y $j = 1, 2, \ldots, n$. En este caso μ se conoce como la **gran media**, y las α_i, llamadas los **efectos del tratamiento**, son tales que $\sum_{i=1}^{k} \alpha_i = 0$. Advierta que hemos escrito meramente la media de la $i$ésima población como $\mu_i = \mu + \alpha_i$ e impuesto la condición $\sum_{i=1}^{k} \alpha_i = 0$ de manera que la media de las μ_i sea igual a la gran media μ. La práctica de referirse a las diferentes poblaciones como diferentes **tratamientos** se debe al hecho que muchas técnicas del análisis de la varianza se desarrollaron originalmente en relación con experimentos agrícolas donde, por ejemplo, diferentes fertilizantes se consideraban como diferentes tratamientos aplicados a la tierra. Así, nos referiremos a los tres detergentes de este ejemplo como tres tratamientos diferentes, y en otros problemas podemos referirnos a cuatro nacionalidades como cuatro tratamientos diferentes, cinco clases de campañas de publicidad como cinco tratamientos diferentes, y así sucesivamente. "Niveles" es otro término que se usa a menudo en vez de "tratamientos".

La hipótesis nula que queremos probar es que las medias de las poblaciones son todas iguales, esto es, que $\mu_1 = \mu_2 = \cdots = \mu_k$ o, equivalentemente, que

$$H_0\colon \quad \alpha_i = 0 \qquad \text{y } 1, 2, \ldots, k$$

Correspondientemente, la hipótesis alternativa es que las medias de las poblaciones no son todas iguales; esto es:

$$H_1: \quad \alpha_i \neq 0 \qquad \text{para al menos un valor de } i$$

La prueba misma se basa en un análisis de la variabilidad total de los datos combinados ($nk - 1$ multiplicado por su varianza), lo cual está dado por

$$\sum_{i=1}^{k} \sum_{j=1}^{n} (x_{ij} - \bar{x}_{\cdot\cdot})^2 \qquad \text{donde} \quad \bar{x}_{\cdot\cdot} = \frac{1}{nk} \cdot \sum_{i=1}^{k} \sum_{j=1}^{n} x_{ij}$$

Si la hipótesis nula es verdadera, toda esta variabilidad se debe al azar, pero si no es verdadera, entonces parte de la suma de los cuadrados anteriores se debe a las diferencias entre las medias de las poblaciones. Para aislar, o separar, estas dos contribuciones a la variabilidad total de los datos, nos referimos al siguiente teorema.

TEOREMA 15.1

$$\sum_{i=1}^{k} \sum_{j=1}^{n} (x_{ij} - \bar{x}_{\cdot\cdot})^2 = n \cdot \sum_{i=1}^{k} (\bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot})^2 + \sum_{i=1}^{k} \sum_{j=1}^{n} (x_{ij} - \bar{x}_{i\cdot})^2$$

donde $\bar{x}_{i\cdot}$ es la media de las observaciones de la *i*ésima población y $\bar{x}_{\cdot\cdot}$ es la media de todas las nk observaciones.

Demostración

$$\begin{aligned}
\sum_{i=1}^{k} \sum_{j=1}^{n} (x_{ij} - \bar{x}_{\cdot\cdot})^2 &= \sum_{i=1}^{k} \sum_{j=1}^{n} [(\bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot}) + (x_{ij} - \bar{x}_{i\cdot})]^2 \\
&= \sum_{i=1}^{k} \sum_{j=1}^{n} [(\bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot})^2 + 2(\bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot})(x_{ij} - \bar{x}_{i\cdot}) \\
&\quad + (x_{ij} - \bar{x}_{i\cdot})^2] \\
&= \sum_{i=1}^{k} \sum_{j=1}^{n} (\bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot})^2 + 2\sum_{i=1}^{k} \sum_{j=1}^{n} (\bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot})(x_{ij} - \bar{x}_{i\cdot}) \\
&\quad + \sum_{i=1}^{k} \sum_{j=1}^{n} (x_{ij} - \bar{x}_{i\cdot})^2 \\
&= n \cdot \sum_{i=1}^{k} (\bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot})^2 + \sum_{i=1}^{k} \sum_{j=1}^{n} (x_{ij} - \bar{x}_{i\cdot})^2
\end{aligned}$$

puesto que $\sum_{j=1}^{n} (x_{ij} - \bar{x}_{i\cdot}) = 0$ para cada valor de i. ▼

Es costumbre referirse a la expresión en el lado izquierdo de la identidad del teorema 15.1 como la suma de **cuadrados total**, al primer término de la expresión en el lado derecho como la **suma de cuadrados de los tratamientos**, y al segundo término como

15.6 Use multiplicadores de Lagrange para demostrar que las estimaciones de mínimos cuadrados de los parámetros del modelo en la página 497 son $\hat{\mu} = \bar{x}_{..}$ y $\hat{\alpha}_i = \bar{x}_{i\cdot} - \bar{x}_{..}$.

APLICACIONES

15.7 Para comparar la eficacia de tres tipos diferentes de recubrimientos fosforescentes para las carátulas de los instrumentos de aviones, se recubren cada una de ocho carátulas con los tres tipos. Entonces se iluminan las carátulas con una luz ultravioleta, y los siguientes son la cantidad de minutos que cada una brilló después de apagar la fuente de luz:

Tipo 1:	52.9, 62.1, 57.4, 50.0, 59.3, 61.2, 60.8, 53.1
Tipo 2:	58.4, 55.0, 59.8, 62.5, 64.7, 59.9, 54.7, 58.4
Tipo 3:	71.3, 66.6, 63.4, 64.7, 75.8, 65.6, 72.9, 67.3

Pruebe la hipótesis nula que no hay diferencia en la eficacia de los tres recubrimientos en el nivel 0.01 de significancia.

15.8 Éste es el número de errores que en cinco semanas sucesivas cometieron cuatro técnicos que trabajan en un laboratorio médico:

Técnico I:	13, 16, 12, 14, 15
Técnico II:	14, 16, 11, 19, 15
Técnico III:	13, 18, 16, 14, 18
Técnico IV:	18, 10, 14, 15, 12

Pruebe en el nivel 0.05 de significancia si las diferencias entre las cuatro muestras se pueden atribuir al azar.

15.9 Tres grupos de seis conejillos de indias se inyectaron, cada uno, con respectivamente 0.5 miligramos, 1.0 miligramo y 1.5 miligramos de un nuevo tranquilizante, y a continuación se muestra el número de minutos que tardaron en quedarse dormidos:

0.5 *mg*:	21, 23, 19, 24, 25, 23
1.0 *mg*:	19, 21, 20, 18, 22, 20
1.5 *mg*:	15, 10, 13, 14, 11, 15

Pruebe en el nivel 0.05 de significancia si se puede rechazar la hipótesis nula de que las diferencias en dosificación no tienen efecto. También, estime los parámetros μ, α_1, α_2 y α_3 del modelo que se usó en el análisis.

15.10 Lo siguiente es el número de palabras por minuto que una secretaria mecanografió en varias ocasiones en cuatro máquinas de escribir diferentes:

Máquina de escribir C:	71, 75, 69, 77, 61, 72, 71, 78
Máquina de escribir D:	68, 71, 74, 66, 69, 67, 70, 62
Máquina de escribir E:	75, 70, 81, 73, 78, 72
Máquina de escribir F:	62, 59, 71, 68, 63, 65, 72, 60, 64

Use las fórmulas de cálculo del ejercicio 15.4 para calcular las sumas de los cuadrados requeridas para probar en el nivel 0.05 de significancia si las diferencias entre las medias de las cuatro muestras se pueden atribuir al azar.

15.11 Un servicio de pruebas para el consumidor, desea probar la exactitud de los termostatos de tres diferentes clases de planchas eléctricas, las puso a 480°F y obtuvo las siguientes lecturas de la temperatura verdadera por medio de un termo par:

Plancha X:	474, 496, 467, 471
Plancha Y:	492, 498
Plancha Z:	460, 495, 490

Use las fórmulas de cálculo del ejercicio 15.4 para calcular las sumas de los cuadrados requeridas para probar en el nivel 0.05 de significancia si las diferencias entre las tres muestras se pueden atribuir al azar.

15.12 En la sección 13.7 señalamos que en el análisis de ji cuadrada de una tabla $r \times c$ no tomamos en consideración un posible orden de los renglones y/o las columnas. Cuando los renglones y las columnas están ambos en orden, indicamos una alternativa al análisis de ji cuadrada en los ejercicios 14.61 y 14.62. Cuando sólo las columnas o sólo los renglones están en orden, consideramos las categorías que no están en orden como tratamientos, y reemplazamos las que están en orden por enteros consecutivos. Por ejemplo, en la tabla de 3×3 en la página 438 consideramos a las tres ciudades como tres tratamientos diferentes, y reemplazamos los encabezados de las columnas con 1, -1 y 0, lo que refleja un orden de favorecer B (no favorecer A) a ser indiferente a favorecer A. Así, la muestra de tamaño $n_1 = 400$ de Los Ángeles consiste de 174 unos, 93 menos unos y 133 ceros; la muestra de tamaño $n_2 = 500$ de San Diego consiste de 196 unos, 124 menos unos y 180 ceros; y así sucesivamente. Al ver la tabla de $r \times c$ de esta manera, podemos entonces realizar un análisis de la varianza en un sentido. Use este método para analizar la tabla de 3×3 de la página 438, pruebe la hipótesis nula de que los efectos del tratamiento son todos igual a cero en el nivel 0.05 de significancia, y compare el resultado con el obtenido en el ejercicio 13.80.

15.13 Use el método del ejercicio 15.12 para analizar la tabla de 3×3 del ejercicio 13.78 y compare el resultado con el resultado obtenido en ese ejercicio.

15.3 DISEÑO DE EXPERIMENTOS

En el ejemplo 15.1 pudiera haber parecido razonable concluir que los tres detergentes no son igualmente eficaces; sin embargo, un momento de reflexión mostrará que esta conclusión no es tan "razonable" después de todo. Realmente no sabemos si las muestras limpiadas con el detergente B podrían haber estado más sucias que las otras, los tiempos de lavado podrían haber sido más largos para el detergente C, podría haber habido diferencias en la dureza o la temperatura del agua, y aun si los instrumentos usados para hacer las lecturas de blancura podrían haberse desajustado después de hacerse las lecturas para los detergentes A y C.

Es totalmente posible, por supuesto, que las diferencias entre las tres medias de las muestras se debieran principalmente a las diferencias en la eficacia de los detergentes, pero hemos enumerado precisamente varios factores que podrían considerarse responsables. Es importante recordar que *una prueba de significancia puede mostrar que las di-*

ferencias entre las medias de las muestras son demasiado grandes para atribuirse al azar, pero esas pruebas no pueden decir por qué ocurrieron las diferencias.

En general, si queremos mostrar que un factor (entre varios más) se puede considerar la causa de un fenómeno observado, debemos, de alguna manera, asegurarnos de que ninguno de los demás factores puede en forma razonable considerarse como responsable. Hay diversas maneras en que esto se puede hacer; por ejemplo, podemos llevar a cabo un **experimento** rigurosamente **controlado** donde todas las variables se mantienen fijas excepto en la que estamos interesados. Para hacer esto en el ejemplo que trata de los tres detergentes, podríamos ensuciar las muestras con exactamente la misma cantidad de tinta china, usar siempre el mismo tiempo de lavado y agua de exactamente la misma dureza y temperatura, e inspeccionar (y si es necesario, ajustar) los instrumentos de medición después de cada uso. Bajo tales condiciones rígidamente controladas, las diferencias significativas entre las medias de las muestras no se pueden deber a muestras que se ensuciaron en forma diferente o a diferencias en tiempos de lavado, temperatura del agua, dureza del agua, o instrumentos de medición. En el lado positivo, las diferencias entre las medias muestran que los detergentes no son todos igualmente eficaces *si se usan en esta forma estrechamente restringida.* Por supuesto, no podemos decir si existirían las mismas diferencias si los tiempos de lavado fueran más largos o más cortos, si el agua tuviera diferente temperatura o dureza, y así sucesivamente.

En la mayoría de los casos, los experimentos "sobrecontrolados" como el que acabamos de describir no proporcionan verdaderamente la clase de información que queremos. Así que buscamos alternativas, y en el otro extremo podemos realizar experimentos donde ninguno de los factores ajenos está controlado, pero en el que nos protegemos contra sus efectos mediante la **aleatorización.** Esto es, diseñamos o planeamos los experimentos de manera que las variaciones causadas por los factores ajenos se puedan combinar todas bajo el encabezado general de "azar". Por ejemplo, podemos conseguir esto al asignar aleatoriamente, en nuestro ejemplo, cinco de las muestras sucias a cada detergente y especificar aleatoriamente el orden en que se lavarán y medirán. Cuando todas las variaciones causadas por factores ajenos no controlados pueden incluirse así bajo el encabezado de variación fortuita, nos referimos al diseño del experimento como **un diseño completamente al azar**.

Sin embargo, debe resultar evidente que la aleatorización protege contra los efectos de factores ajenos sólo en una manera probabilística. Por ejemplo, en nuestro ejemplo es posible, aunque muy poco probable, que el detergente A sea aleatoriamente asignado a las cinco muestras que resultan ser las menos sucias o que el agua resulta ser la más fría cuando lavamos las cinco muestras del detergente B. Es en parte por esta razón que a menudo tratamos de controlar algunos de los factores y dejar al azar otros, y así usamos diseños que están en algún punto entre los dos extremos que hemos descrito.

Para introducir otro concepto importante en el diseño de experimentos, consideremos los datos sobre la cantidad de tiempo (en minutos) que tomó a cierta persona conducir hasta su trabajo, de lunes a viernes, por cuatro rutas diferentes:

Ruta 1:	22, 26, 25, 25, 31
Ruta 2:	25, 27, 28, 26, 29
Ruta 3:	26, 29, 33, 30, 33
Ruta 4:	26, 28, 27, 30, 30

Las medias de estas cuatro muestras son 25.8, 27.0, 30.2 y 28.2, y puesto que las diferencias son bastante grandes, parecería razonable concluir que hay algunas diferencias verdaderas en los promedios verdaderos del tiempo que tarda la persona en conducir al trabajo por cuatro rutas diferentes. Esto no se sigue, sin embargo, de un análisis de la varianza en un sentido. Obtenemos $f = 2.80$; puesto que esto no excede a $f_{0.05, 3, 16} = 3.24$, no se puede rechazar la hipótesis nula.

Por supuesto, la hipótesis nula puede ser verdadera, pero observe que no sólo hay diferencias considerables entre las cuatro medias, sino también diferencias grandes entre los valores dentro de las muestras. En la primera muestra varían de 22 a 31, en la segunda muestra de 25 a 29, en la tercera muestra de 26 a 33 y en la cuarta muestra de 26 a 30. Y no sólo eso, sino que en cada muestra el primer valor es el más pequeño y el último valor el más grande. Esto último sugiere que la variación dentro de las muestras bien puede deberse a las diferencias en las condiciones de manejo en los diferentes días de la semana. Si éste es el caso, las variaciones causadas por las condiciones de manejo se incluyeron en la suma de cuadrados del error del análisis de la varianza en un sentido, se "infló" el denominador de la estadística f y éste puede ser la razón por qué los resultados no fueron significativos.

Para evitar esta clase de situación, podríamos mantener fijos los factores ajenos, pero esto rara vez nos dará la información que queremos. En nuestro ejemplo, podríamos limitar el estudio a las condiciones de manejo del lunes, pero entonces no tendríamos la seguridad de que los resultados se aplicarían también a las condiciones de manejo de los martes o de cualquier otro día de la semana. Otra posibilidad es variar el factor ajeno deliberadamente en un intervalo tan amplio como sea necesario de manera que la variación que causa se pueda medir, y, por tanto, eliminar de la suma de cuadrados del error. Esto significa que debemos planear el experimento de manera que podamos realizar un **análisis de la varianza en dos sentidos**, en el que la variación total de los datos se divide en tres componentes atribuidos, respectivamente, a los tratamientos (en nuestro ejemplo, las rutas), el factor ajeno (en nuestro caso, las condiciones de manejo en los diferentes días de la semana), y el error experimental, o azar.

Lo que hemos sugerido aquí se llama **conformación de bloques**, y los diferentes días de la semana se conocen como **bloques**. En general, los bloques son los niveles en que mantenemos fijo un factor ajeno de manera que podamos medir su contribución a la variación total de los datos. Si cada tratamiento aparece el mismo número de veces en cada bloque (en nuestro ejemplo, cada ruta se usa una vez de cada día de la semana), decimos que el diseño del experimento es un **diseño de bloque completo**. Además, si los tratamientos se distribuyen aleatoriamente dentro de cada bloque (en nuestro ejemplo, distribuiríamos aleatoriamente las cuatro rutas entre los cuatro lunes, los cuatro martes, y así sucesivamente), decimos que el diseño del experimento es **un diseño en bloques al azar**.

15.4 ANÁLISIS DE LA VARIANZA EN DOS SENTIDOS SIN INTERACCIÓN

Hay esencialmente dos formas diferentes de analizar los experimentos de dos variables, y depende de si las dos variables son independientes o si **interaccionan.** Para ilustrar lo que queremos decir aquí por "interaccionan", suponga que una fabricante de neumáticos está experimentando con diferentes neumáticos y encuentra que una clase es especialmente buena en carreteras de terracería, mientras que otra clase es especialmente buena para uso en pavimento duro. Si éste es el caso, decimos que hay una **interacción** entre las condiciones de la carretera y el diseño del neumático. Primero, sólo estudiaremos el caso de no interacción y después abordaremos el caso de interacción en la sección 15.5.

Para presentar la teoría del análisis de la varianza en dos sentidos, usaremos la terminología introducida en las secciones precedentes y nos referiremos a los dos variables como tratamientos y bloques; en forma alternativa, también nos podemos referir a ellos como el **factor A** y el **factor B** o como **renglones** y **columnas**. Así, si x_{ij} para $i = 1, 2, \ldots, k$ y $j = 1, 2, \ldots, n$ son los valores de variables aleatorias independientes que tienen distribuciones normales con las respectivas medias μ_{ij} y la varianza común σ^2, consideraremos el arreglo

	Bloque 1	*Bloque* 2	...	*Bloque n*
Tratamiento 1	x_{11}	x_{12}	...	x_{1n}
Tratamiento 2	x_{21}	x_{22}	...	x_{2n}
...	...	...	...	...
Tratamiento k	x_{k1}	x_{k2}	...	x_{kn}

y escribimos el modelo para un análisis de la varianza en dos sentidos (sin interacción) como

$$x_{ij} = \mu + \alpha_i + \beta_j + e_{ij}$$

para $i = 1, 2, \ldots, k$ y $j = 1, 2, \ldots, n$. En este caso μ es la **gran media**, los **efectos del tratamiento** α_i son tales que $\sum_{i=1}^{k} \alpha_i = 0$, los **efectos de bloque** β_j son tales que $\sum_{j=1}^{n} \beta_j = 0$, y las e_{ij} son valores de variables aleatorias independientes que tienen distribuciones normales con media cero y la varianza común σ^2. Observe que

$$\mu_{ij} = \mu + \alpha_i + \beta_j$$

y, como se pedirá al lector que lo verifique en el ejercicio 15.15,

$$\frac{\sum_{i=1}^{k} \sum_{j=1}^{n} \mu_{ij}}{nk} = \mu$$

Las dos hipótesis nulas que queremos probar son que los efectos del tratamiento son todos igual a cero y que los efectos de los bloques son todos igual a cero; esto es

$$H_0\colon \quad \alpha_i = 0 \qquad \text{para } i = 1, 2, \ldots, k$$

y

$$H_0'\colon \quad \beta_j = 0 \qquad \text{para } j = 1, 2, \ldots, n$$

La alternativa a H_0 es que los efectos del tratamiento no son todos iguales a cero, y la alternativa a H_0' es que los efectos de los bloques no son todos iguales a cero. Simbólicamente,

$$H_1\colon \quad \alpha_i \neq 0 \qquad \text{para al menos un valor de } i$$

y

$$f_B = \frac{\dfrac{\text{SSB}}{(n-1)\sigma^2}}{\dfrac{\text{SSE}}{(n-1)(k-1)\sigma^2}} = \frac{\text{MSB}}{\text{MSE}}$$

Esta clase de análisis se llama un **análisis de la varianza en dos sentidos**, y los detalles necesarios suelen presentarse en el siguiente tipo de tabla de análisis de la varianza:

Fuente de variación	*Grados de libertad*	*Suma de cuadrados*	*Cuadrado medio*	f
Tratamientos	$k-1$	SS(Tr)	MS(Tr)	$f_{\text{Tr}} = \frac{\text{MS(Tr)}}{\text{MSE}}$
Bloques	$n-1$	SSB	MSB	$f_B = \frac{\text{MSB}}{\text{MSE}}$
Error	$(n-1)(k-1)$	SSE	MSE	
Total	$nk-1$	SST		

Para simplificar los cálculos, SST y SS(Tr) suelen determinarse por medio de las fórmulas del teorema 15.2, y SSB se puede determinar por medio de la fórmula siguiente, la cual se pedirá al lector que la derive en el ejercicio 15.17.

TEOREMA 15.4

$$\text{SSB} = \frac{1}{k}\cdot\sum_{j=1}^{n} T_{\cdot j}^2 - \frac{1}{kn}\cdot T_{\cdot\cdot}^2$$

donde $T_{\cdot j}$ es el total de los valores obtenidos en el $j$ésimo bloque y $T_{\cdot\cdot}$ es el total general de las nk observaciones.

Entonces, el valor de SSE se puede obtener al restar SS(Tr) y SSB de SST.

EJEMPLO 15.2

Con respecto a la ilustración en la página 505, donde teníamos

	Lunes	*Martes*	*Miércoles*	*Jueves*	*Viernes*
Ruta 1	22	26	25	25	31
Ruta 2	25	27	28	26	29
Ruta 3	26	29	33	30	33
Ruta 4	26	28	27	30	30

pruebe en el nivel 0.05 de significancia si las diferencias entre las medias obtenidas por las diferentes rutas (tratamientos) son significativas y también si las diferencias entre las medias obtenidas para los diferentes días de la semana (bloques) son significativas.

Solución

1. H_0: $\alpha_i = 0$ para $i = 1, 2, 3, 4$
   H_0': $\beta_j = 0$ para $j = 1, 2, 3, 4, 5$
   H_1: $\alpha_i \neq 0$ para al menos un valor de i
   H_1': $\beta_j \neq 0$ para al menos un valor de j.
   $\alpha = 0.05$ para ambas pruebas.
2. Rechace la hipótesis nula para los tratamientos si $f_{Tr} \geqq 3.49$ y rechace la hipótesis nula para los bloques si $f_B \geqq 3.26$, donde f_{Tr} y f_B se obtienen por medio de un análisis de la varianza en dos sentidos, y 3.49 y 3.26 son, respectivamente, los valores de $f_{0.05,3,12}$ y $f_{0.05,4,12}$.
3. Las sumas y sumas de los cuadrados requeridas son $T_1. = 129$, $T_2. = 135$, $T_3. = 151$, $T_4. = 141$, $T._1 = 99$, $T._2 = 110$, $T._3 = 113$, $T._4 = 111$, $T._5 = 123$, $T.. = 556$ y $\sum\sum x^2 = 15{,}610$, y la sustitución de estos valores junto con $k = 4$ y $n = 5$ en las fórmulas del teorema 15.2 y 15.4 nos da

$$\begin{aligned} \text{SST} &= 15{,}610 - \frac{1}{20}(556)^2 \\ &= 153.2 \end{aligned}$$

$$\begin{aligned} \text{SS(Tr)} &= \frac{1}{5}(129^2 + 135^2 + 151^2 + 141^2) - \frac{1}{20}(556)^2 \\ &= 52.8 \end{aligned}$$

$$\begin{aligned} \text{SSB} &= \frac{1}{4}(99^2 + 110^2 + 113^2 + 111^2 + 123^2) - \frac{1}{20}(556)^2 \\ &= 73.2 \end{aligned}$$

y por tanto

$$\begin{aligned} \text{SSE} &= 153.2 - 52.8 - 73.2 \\ &= 27.2 \end{aligned}$$

Los cálculos restantes se muestran en la siguiente tabla de análisis de la varianza:

Fuente de variación	*Grados de libertad*	*Suma de cuadrados*	*Cuadrado medio*	f
Tratamientos	3	52.8	$\frac{52.8}{3} = 17.6$	$\frac{17.6}{2.27} = 7.75$
Bloques	4	73.2	$\frac{73.2}{4} = 18.3$	$\frac{18.3}{2.27} = 8.06$
Error	12	27.2	$\frac{27.2}{12} = 2.27$	
Total	19	153.2		

4. Puesto que $f_{Tr} = 7.75$ excede a 3.49 y $f_B = 8.06$ excede a 3.26, se deben rechazar ambas hipótesis nulas. En otras palabras, las diferencias entre las medias obtenidas para las cuatro rutas son significativas y también lo son las diferencias entre las medias obtenidas para los diferentes días de la semana. Sin embargo, advierta que no podemos concluir que la ruta 1 es necesariamente la más rápida y que en viernes las condiciones de tráfico son siempre las peores. Todo lo que hemos mostrado por medio del análisis es que las diferencias existen, y si queremos ir un paso más allá y precisar la naturaleza de las diferencias, tendremos que usar una **prueba de comparaciones múltiple** tal como la de la sección 15.6. ▲

EJERCICIOS

15.14 Haga uso de la identidad

$$x_{ij} - \bar{x}_{..} = (\bar{x}_{i.} - \bar{x}_{..}) + (\bar{x}_{.j} - \bar{x}_{..}) + (x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{..})$$

para demostrar el teorema 15.3.

15.15 Con respecto a la notación de la página 508, muestre que

$$\frac{\sum_{i=1}^{k} \sum_{j=1}^{n} \mu_{ij}}{nk} = \mu$$

15.16 Para el análisis de la varianza en dos sentidos con k tratamientos y n bloques, muestre que

$$E\left[\frac{k \cdot \sum_{j=1}^{n} (\bar{X}_{.j} - \bar{X}_{..})^2}{n-1}\right] = \sigma^2 + \frac{k \cdot \sum_{j=1}^{n} \beta_j^2}{n-1}$$

15.17 Demuestre el teorema 15.4.

15.18 Un **cuadrado latino** es un arreglo cuadrado donde cada letra (o cierta clase de símbolo) aparece exactamente una vez en cada renglón y una vez en cada columna. Por ejemplo,

A	*B*	*C*	*D*
B	*C*	*D*	*A*
C	*D*	*A*	*B*
D	*A*	*B*	*C*

es un cuadrado latino 4×4 Si consideramos los m renglones de un cuadrado latino como los niveles de una variable, las m columnas como los niveles de una segunda variable, y $A, B, C, \ldots$, como m "tratamientos", esto es, como los niveles de una tercera variable, es posible probar las hipótesis concernientes a todas estas tres variables con base en tan pocas observaciones como m^2 (siempre y cuando no haya interacciones). Sea que $x_{ij(k)}$ al denotar la observación en el *iésimo* renglón y la *j*ésima columna de un cuadrado latino (de manera que k, que denota el tratamiento, se determina al especificar i y j), escribimos la ecuación modelo como

$$x_{ij(k)} = \mu + \alpha_i + \beta_j + \tau_k + e_{ij}$$

para $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, m$, y $k = 2, \ldots, m$, donde μ es la gran media, los **efectos de los renglones** α_i son tales que $\sum_{i=1}^{m} \alpha_i = 0$, los **efectos de las columnas** β_j son tales que $\sum_{j=1}^{m} \beta_j = 0$, los efectos de los tratamientos τ_k son tales que $\sum_{k=1}^{m} \tau_k = 0$, y las e_{ij} son los valores de variables aleatorias independientes que tienen distribuciones normales con media cero y la varianza común σ^2. La hipótesis nula que queremos probar (contra alternativas apropiadas) es que los efectos de los renglones son todos cero, que los efectos de las columnas son todos cero, y que los efectos de los tratamientos son todos cero.

(a) Muestre que

$$\sum_{i=1}^{m} \sum_{j=1}^{m} (x_{ij(k)} - \bar{x}_{..})^2 = m \cdot \sum_{i=1}^{m} (\bar{x}_{i.} - \bar{x}_{..})^2 + m \cdot \sum_{j=1}^{m} (\bar{x}_{.j} - \bar{x}_{..})^2$$

$$+ m \cdot \sum_{k=1}^{m} (\bar{x}_{(k)} - \bar{x}_{..})^2 + \sum_{i=1}^{m} \sum_{j=1}^{m} (x_{ij(k)} - \bar{x}_{i.} - \bar{x}_{.j} - \bar{x}_{(k)} + 2\bar{x}_{..})^2$$

donde $\bar{x}_{(k)}$ es la media de todas las observaciones para el *k*ésimo tratamiento y las otras medias son como se define en el teorema 15.3. La expresión en el lado izquierdo de la identidad anterior es la suma de cuadrados total SST, mientras que las del lado derecho son, respectivamente, la **suma de cuadrados de los renglones** SSR, la **suma de cuadrados de las columnas** SSC, la suma de cuadrados de los tratamientos SS(Tr), y la suma de los cuadrados del error SSE.

(b) Construya una tabla de análisis de la varianza para esta clase de experimento, determine los grados de libertad para SSE al restar los de SSR, SSC y SS(Tr) de $m^2 - 1$, los grados de libertad de SST.

APLICACIONES

15.19 Se lleva a cabo un experimento para juzgar los efectos de cuatro diferentes combustibles y tres tipos diferentes de lanzador sobre el alcance de cierto cohete. Pruebe, con base en los siguientes alcances en millas, si hay un efecto significativo a causa de la diferencia en combustibles y si hay un efecto significativo a causa de las diferencias en lanzadores:

	Combustible 1	*Combustible* 2	*Combustible* 3	*Combustible* 4
Lanzador X	45.9	57.6	52.2	41.7
Lanzador Y	46.0	51.0	50.1	38.8
Lanzador Z	45.7	56.9	55.3	48.1

Use el nivel 0.01 de significancia.

15.20 Los siguientes son los contenidos de colesterol en miligramos por paquete que obtuvieron cuatro laboratorios para paquetes de 6 onzas de tres alimentos dietéticos muy similares:

	Alimento dietético A	*Alimento dietético B*	*Alimento dietético C*
Laboratorio 1	3.4	2.6	2.8
Laboratorio 2	3.0	2.7	3.1
Laboratorio 3	3.3	3.0	3.4
Laboratorio 4	3.5	3.1	3.7

Realice un análisis de la varianza en dos sentidos y pruebe las hipótesis nulas concernientes a los alimentos dietéticos y los laboratorios en el nivel 0.05 de significancia.

15.21 Un técnico de laboratorio mide la resistencia a la ruptura de cada una de cinco clases de hilo de lino utilizando cuatro diferentes instrumentos de medición, I_1, I_2, I_3 e I_4, y obtiene los resultados siguientes, en onzas:

	I_1	I_2	I_3	I_4
Hilo 1	20.9	20.4	19.9	21.9
Hilo 2	25.0	26.2	27.0	24.8
Hilo 3	25.5	23.1	21.5	24.4
Hilo 4	24.8	21.2	23.5	25.7
Hilo 5	19.6	21.2	22.1	22.1

Realice un análisis de la varianza en dos sentidos, usando el nivel 0.05 de significancia.

15.22 Los datos muestrales en el siguiente cuadrado latino (véase el ejercicio 15.18) son las puntuaciones obtenidas por ocho estudiantes universitarios de diversos orígenes étnicos y diversos intereses profesionales en una prueba de historia de Estados Unidos:

donde $\bar{x}_{i..}$ es la media de las observaciones del $i$ésimo valor del primer tratamiento, $\bar{x}_{.j.}$ es la media de el $j$ésimo valor del segundo tratamiento, $\bar{x}_{..r}$ es la media de la $r$ésima réplica, $\bar{x}_{ij.}$ es la media del $i$ésimo y $j$ésimo valores de los dos tratamientos (promediados sobre las réplicas) y $\bar{x}_{...}$ es la gran media de todas las rnm observaciones.

Demostración. Para probar el teorema, primero escribimos la identidad

$$\begin{aligned} x_{ijr} - \bar{x}_{...} &= (\bar{x}_{i..} - \bar{x}_{...}) + (\bar{x}_{.j.} - \bar{x}_{...}) + (\bar{x}_{..r} - \bar{x}_{...}) \\ &\quad + (\bar{x}_{ij.} - \bar{x}_{i..} - \bar{x}_{.j.} + \bar{x}_{...}) + (\bar{x}_{ijk} - \bar{x}_{ij.} - \bar{x}_{..r} + \bar{x}_{...}) \end{aligned}$$

Cuando elevamos al cuadrado cada lado de esta identidad y sumamos sobre i, j y r, se puede mostrar que todos los términos con productos cruzados suman cero. Los detalles de la demostración de este teorema se dejan al lector en el ejercicio 15.25. ▼

Análoga a la clasificación en dos sentidos sin interacción, la expresión en el lado izquierdo de la identidad del teorema 15.5 es la suma de cuadrados total, SST, y los dos primeros términos en la derecha son la suma de cuadrados de los tratamientos, que ahora denotaremos con SSA y SSB. El tercer término en el lado derecho es la suma de cuadrados para las réplicas, SSR, el cuarto término es la suma de cuadrados para las interacciones, SSI, y el término final es la *nueva* suma de cuadrados de los errores, SSE. Así,

$$\text{SST} = \text{SSA} + \text{SSB} + \text{SSR} + \text{SSI} + \text{SSE}$$

y se puede mostrar que si $H_0^{(1)}, \ldots, H_0^{(4)}$ son verdad, las cantidades

$$f_A = \frac{\dfrac{\text{SSA}}{(k-1)\sigma^2}}{\dfrac{\text{SSE}}{(m-1)(nk-1)\sigma^2}} = \frac{\text{MSA}}{\text{MSE}}$$

$$f_B = \frac{\dfrac{\text{SSB}}{(n-1)\sigma^2}}{\dfrac{\text{SSE}}{(m-1)(nk-1)\sigma^2}} = \frac{\text{MSB}}{\text{MSE}}$$

$$f_R = \frac{\dfrac{\text{SSR}}{(m-1)\sigma^2}}{\dfrac{\text{SSE}}{(m-1)(nk-1)\sigma^2}} = \frac{\text{MSR}}{\text{MSE}}$$

$$f_I = \frac{\dfrac{\text{SSI}}{(n-1)(k-1)\sigma^2}}{\dfrac{\text{SSE}}{(m-1)(nk-1)\sigma^2}} = \frac{\text{MSI}}{\text{MSE}}$$

todas tienen distribuciones F con, respectivamente, $k-1$, $n-1$, $m-1$ y $(k-1)(n-1)$ grados de libertad en el numerador y $(m-1)(nk-1)$ grados de libertad en

Diseños

	A	*B*	*C*	*D*	*Totales*
Noreste	107	59	132	125	423
Sureste	78	40	118	88	324
Noroeste	122	60	141	100	423
Suroeste	65	22	99	71	257
Totales	372	181	490	384	1427

Así, por ejemplo, $T_{1\cdot\cdot} = 372$, $T_{\cdot 2\cdot} = 324$, $T_{11\cdot} = 107$ y así sucesivamente. También, calculamos a partir de los datos originales $T_{\cdot\cdot 1} = 738$ y $T_{\cdot\cdot 2} = 689$. La suma de cuadrados total es $\sum \sum \sum x^2 = 73{,}667$. La sustitución de estos valores junto con $k = n = 4$ y $r = 2$ en las fórmulas del teorema 15.6 nos da

$$C = \frac{1}{32}(1{,}427)^2 = 63{,}635$$

al entero más cercano y

$$\text{SSA} = \frac{1}{8}(372^2 + 181^2 + 490^2 + 384^2) - 63{,}635$$

$$= 6{,}203$$

$$\text{SSB} = \frac{1}{8}(423^2 + 324^2 + 423^2 + 257^2) - 63{,}635$$

$$= 2{,}475$$

$$\text{SSR} = \frac{1}{16}(738^2 + 689^2) - 63{,}635$$

$$= 75$$

$$\text{SSI} = \frac{1}{2}(107^2 + 59^2 + 132^2 + \cdots + 99^2 + 71^2) - 6{,}203 - 2{,}475 - 63{,}635$$

$$= 311$$

y por tanto

$$\text{SSE} = 73{,}667 - 6{,}203 - 2{,}475 - 75 - 311 - 63{,}635$$

$$= 968$$

Los cálculos restantes se muestran en la siguiente tabla de análisis de la varianza:

	Combustible 1	*Combustible 2*	*Combustible 3*	*Combustible 4*
Lanzador X	46.1	55.9	52.6	44.3
Lanzador Y	46.3	52.1	51.4	39.6
Lanzador Z	45.8	57.9	56.2	47.6

Combine estos datos con los del ejercicio 15.19, realice un análisis de la varianza apropiado para probar la hipótesis nula que incluya combustibles, lanzadores, réplicas y la interacción combustible con lanzador. Use el nivel 0.01 de significancia.

15.28 El experimento descrito en el ejercicio 15.20 se repitió con los resultados siguientes.

	Alimento dietético A	*Alimento dietético B*	*Alimento dietético C*
Laboratorio 1	3.5	2.5	2.9
Laboratorio 2	3.0	2.9	3.2
Laboratorio 3	3.6	3.4	3.8
Laboratorio 4	3.3	3.5	3.4

Combine estos datos con los del ejercicio 15.20, realice un análisis de la varianza apropiado para probar la hipótesis nula que involucre alimentos dietéticos, laboratorios, réplicas, y la interacción de alimentos con laboratorio. Use el nivel 0.05 de significancia.

15.29 Tres operadores operan cada uno cuatro máquinas de interconexión diferentes (usadas para conectar eléctricamente los alambres delgados en la fabricación de circuitos integrados). Los operadores se asignaron aleatoriamente a las máquinas de interconexión. Después se repitió el experimento con una nueva aleatorización de los operadores a las máquinas de interconexión. Después de que se hicieron todas las interconexiones, se probó la resistencia de la conexión de cada una midiendo el número de gramos de fuerza requerida para romper la conexión. Los resultados de los experimentos son como sigue.

	Réplica 1				*Réplica* 2			
Máquina de interconexión	*A*	*B*	*C*	*D*	*A*	*B*	*C*	*D*
Operador 1	11.8	9.6	12.6	10.2	10.6	11.9	9.8	9.9
Operador 2	10.4	12.4	11.0	10.5	12.0	10.3	10.0	11.6
Operador 3	9.6	10.2	11.4	3.1	11.8	9.9	9.1	5.8

Realice un análisis de la varianza apropiado para probar la hipótesis nula concerniente a operadores, máquinas de interconexión, réplicas y la interacción del operador con la máquina de interconexión en el nivel 0.05 de significancia.

15.30 Se usó un índice de sabores para evaluar el efecto de añadir dioctil sulfosuccinato de sodio (DSS) a la leche para estabilizar su sabor. Se usaron cuatro niveles de DSS

2. La tabla IX da valores de r_p para los niveles de significancia de 0.05 y 0.01, dependiendo del número de grados de libertad para el error en el análisis de la varianza y de p, el número de medias que se está comparando.
3. Calcule **el intervalo de significancia mínimo,** use la fórmula

$$R_p = r_p \cdot s_{\bar{x}}$$

4. Ordene las medias por tamaño, de la más pequeña a la más grande.
5. Compare la diferencia de la última y la primera media con R_k. Si esta diferencia es mayor que R_k, se puede concluir que las k medias muestrales son significativamente diferentes con el nivel de significancia usado para determinar r_k de la tabla IX. En la misma forma, compare todos los conjuntos adyacentes de $k - 1$ medias, use ahora R_{k-1} como el criterio de significancia. Continúe este procedimiento para los conjuntos de $k - 2$ medias adyacentes, y así sucesivamente, hasta llegar a conjuntos de dos medias adyacentes. Al hacer estas comparaciones, es útil subrayar las medias adyacentes en un conjunto cuyas medias no son significativamente diferentes. Si entre la comparaciones posteriores hay un subconjunto de medias ya conectadas por un subrayado, no se necesita hacer comparaciones adicionales entre las medias en ese subconjunto.

EJEMPLO 15.4

Con respecto al ejemplo 15.2, use la prueba de intervalo múltiple de Duncan en el nivel 0.05 de significancia para determinar la naturaleza de las diferencias entre las medias de los tratamientos.

Solución

1. De la tabla de análisis de la varianza en la página 511, tenemos MSE = 2.27; así,

$$s_{\bar{x}} = \sqrt{\frac{2.27}{5}} = 0.67$$

2. De la tabla IX con $\alpha + 0.05$ y 12 grados de libertad, obtenemos los siguientes valores de r_p.

p	2	3	4
r_p	3.08	3.23	3.31

3. Al multiplicar cada valor de r_p por $s_{\bar{x}} = 0.67$, obtenemos

p	2	3	4
R_p	2.06	2.16	2.22

4. A continuación ordenamos las cuatro medias de acuerdo a su tamaño, como sigue

Ruta	1	2	4	3
Media	25.8	27.0	28.2	30.2

5. La diferencia entre la media más grande y la más pequeña es $30.2 - 25.8 = 4.4$, lo cual excede a 2.22, el valor de R_4. Así, ningún subrayado conecta a las cuatro medias. (Este resultado era esperado, como el análisis de la va-

rianza mostró una diferencia significativa entre las cuatro medias en el nivel 0.05 de significancia.) Al comparar la diferencia entre la media más grande y la segunda media más pequeña, obtenemos $30.2 - 27.0 = 3.2$, lo cual excede a $R_3 = 2.16$, y al comparar el otro conjunto de tres medias adyacentes, obtenemos $28.2 - 25.8 = 2.4$, lo cual también excede a 2.16. A continuación, al comparar la media más grande con la segunda media más grande $30.2 - 28.2 = 2.0$, lo cual no excede a R_2 2.06. Así, estas dos medias no son significativamente diferentes, y se pueden conectar mediante un subrayado. De la misma manera, al comparar los otros dos juegos de las dos medias adyacentes, obtenemos $28.2 - 27.0 = 1.2$ y $27.0 - 25.8 = 1.2$. Así, podemos conectar estos pares de medias con un subrayado, obteniendo finalmente

Ruta	1	2	4	3
Media	25.8	27.0	28.2	30.2

▲

Enunciando el resultado mostrado en el ejemplo 15.4 con palabras, podemos decir que las rutas 1 y 2 no están asociadas con tiempos de manejo estadísticamente diferentes, pero *como un grupo* tienen tiempos de manejo significativamente diferentes que las otras dos rutas en el nivel 0.05 de significancia. En la misma forma, las rutas 2 y 3 no son "significativamente diferentes", pero *como un grupo* tienen tiempos de manejo significativamente más grande que el primer grupo y tiempos de manejo significativamente más pequeños que el último grupo.

Este resultado tal vez no sea tan definitivo como nos gustaría (por ejemplo, la ruta 2 aparece en ambos grupos, el más bajo y el de en medio). Sin embargo, se pueden tomar decisiones razonables con base en la prueba. Por ejemplo, sería racional escoger la ruta 1 o la ruta 2 si el objetivo es minimizar el tiempo de manejo. Uno podría escoger entre estas rutas con base en la seguridad, el paisaje, o algún otro criterio adicional.

Sin embargo, con este objetivo en mente, no sería razonable escoger la ruta 3 o la ruta 4.

APLICACIONES

15.31 Realice una prueba de intervalos múltiples para determinar la naturaleza de las diferencias entre los tres detergentes en el ejemplo 15.1. Use el nivel 0.01 de significancia.

15.32 Realice una prueba de intervalos múltiples para determinar la naturaleza de las diferencias de bloque en el ejemplo 15.2. Use el nivel 0.05 de significancia.

15.33 Realice una prueba de intervalos múltiples para caracterizar las diferencias entre los diseños de compresores y entre las regiones en el ejemplo 15.3. Use el nivel 0.05 de significancia.

15.34 Realice pruebas apropiadas de amplitud múltiple, use el nivel 0.05 de significancia, para caracterizar las diferencias entre las medias de los alimentos dietéticos y las medias de los laboratorios en el ejercicio 15.28. ¿Bajo qué circunstancias no sería apropiado hacer una prueba como ésa?

15.35 Realice pruebas apropiadas de intervalos múltiples, use el nivel 0.01 de significancia, para caracterizar las diferencias entre las medias de los lanzadores y las

CAPÍTULO

16

Pruebas no paramétricas

16.1 INTRODUCCIÓN

En el capítulo 10 introdujimos el concepto de **robustez** en relación con problemas de estimación. Extendamos ahora este concepto a las pruebas de hipótesis, que se dice son robustas si las distribuciones muestrales de las estadísticas de prueba no están seriamente afectadas por violaciones en las suposiciones sustentantes.

En relación con las pruebas de hipótesis, es especialmente importante saber si las violaciones de las suposiciones sustentantes pudieran afectar el nivel de significancia. Como vimos en la sección 12.5, cualquier comparación de las funciones de potencia de dos o más pruebas requiere que los niveles de significancia sean iguales; y si éste no es el caso, la comparación es inválida. Por ejemplo, la prueba t de una muestra de la sección 13.3 requiere que nuestra muestra venga de una población normal. Así, ¿qué pasa cuando la población es "no muy normal", digamos, si tiene forma de campana pero no es perfectamente simétrica? Las simulaciones con computadoras han mostrado que aun cuando una población puede apartarse algo de la normalidad, la mayor parte del tiempo el nivel de significancia todavía estará cercano a los valores prescritos de α.

Los siguientes ejemplos muestran cómo la violación de las suposiciones sustentantes acerca de una población puede afectar el nivel de significancia. Supongamos que queremos probar la hipótesis nula $\mu = \mu_0$ en el nivel 0.05 de significancia, donde μ es la media de una población normal con desviación estándar *conocida* σ, pero hay una probabilidad importante (digamos, una en 50) de que uno de los valores se registrará incorrectamente. En relación con la prueba que se ilustra en el ejemplo 13.1, estamos así violando la suposición de que estamos tratando con una muestra aleatoria de una población normal. Si uno de los valores en el ejemplo 13.1 se hubiese registrado incorrectamente, digamos, como 7.452 onzas en vez de 7.952 onzas, la media del peso de los 25 paquetes de galletas se habría visto reducido por

$$\frac{7.952 - 7.452}{25} = 0.020$$

onzas, z se hubiese visto reducida de 2.48 a 2.22, y el valor P correspondiente hubiese aumentado de 0.0046 a 0.0264. Puesto que el nuevo valor de P excede a 0.025, ya no se puede rechazar la hipótesis nula; esto muestra cómo los valores P y por tanto el nivel de significancia pueden verse afectados cuando permitimos la posibilidad de registrar incorrectamente los datos.

Ahora supongamos que en un problema como el anterior σ es *desconocida*, de manera que el procedimiento estándar sería la prueba t de una muestra ilustrada en el ejemplo 13.3. En ese caso, un error al registrar un valor afectará la desviación estándar muestral así como la media muestral, que aparecen, respectivamente, en el denominador y numerador de la estadística de prueba. Como se ilustra para un caso especial en el ejercicio 16.1, esto a menudo dará valores de t más cercanos a $+1$ o -1 y, por tanto, hará más difícil rechazar la hipótesis nula. En otras palabras, con el riesgo de un error así, el nivel de significancia bien puede ser menor que el valor α prescrito. Esto también se ilustra en los ejercicios 13.17 y 13.18 de la página 423.

Puesto que hay muchas situaciones donde nos enfrentamos a dudas serias sobre la robustez de las pruebas de hipótesis, en especial con relación a las suposiciones de normalidad, los estadísticos han desarrollado técnicas alternativas que requieren menos suposiciones, si es que requieren alguna. Estas pruebas en general se conocen como **no paramétricas**; entre ellas se encuentran pruebas que están **libres de distribución** (donde no hacemos suposiciones sobre la población, excepto, quizá, que son continuas) y también pruebas que son no paramétricas sólo en cuanto a que no estamos preocupados por los parámetros específicos de las poblaciones dadas.

Aparte del hecho que las pruebas no paramétricas se pueden usar en condiciones más generales que las pruebas estándar a las que reemplazan, tienen una atracción intuitiva considerable, por lo general, son fáciles de explicar y fáciles de entender. También, en muchas pruebas no paramétricas la carga de cálculo es tan ligera que pueden caer bajo el encabezado de técnicas "rápido y fácil" o "atajos". En parte por estas razones, las pruebas no paramétricas se han vuelto muy populares, y hay una extensa literatura dedicada a su teoría y aplicación.

La principal desventaja de las pruebas no paramétricas es que a menudo desperdician la información y así son menos eficientes que las técnicas estándar que reemplazan. Se debe observar, sin embargo, que las comparaciones de eficiencia suelen suponer que se satisfacen las condiciones que sustentan a las pruebas estándar, y por lo tanto tienden a subestimar el valor real de los métodos no paramétricos cuando se tratan los asuntos de robustez. En general, es verdad que *cuanto menos se suponga, tanto menos puede inferirse de un conjunto de datos*; pero también es verdad que *cuanto menos se suponga, tanto más se amplía la aplicabilidad de nuestro método.*

16.2 LA PRUEBA DEL SIGNO

La **prueba del signo** se usa a menudo como una alternativa no paramétrica a la prueba t de una muestra, donde probamos la hipótesis nula $\mu = \mu_0$ contra una alternativa apropiada. Para la prueba del signo, suponemos meramente que la población muestreada es continua y simétrica. Suponemos que la población es continua de manera que hay cero probabilidad de obtener un valor igual a μ_0, y ni siquiera necesitamos la suposición de simetría si cambiamos la hipótesis nula a $\tilde{\mu} = \tilde{\mu}_0$, donde $\tilde{\mu}$ es la mediana de la población.

En la prueba del signo reemplazamos cada valor de la muestra que exceda a μ_0 con un signo más y con cada valor menor que μ_0 con un signo menos, y después probamos la hipótesis nula que el número de signos positivos es un valor de una variable aleatoria que tiene la distribución binomial con los parámetros n (el número total de signos positivos o negativos) y $\theta = \frac{1}{2}$. La alternativa bilateral $\mu \neq \mu_0$ se vuelve así $\theta \neq \frac{1}{2}$, y las alternativas unilaterales $\mu \rightarrow \mu_0$ y $\mu > \mu_0$ se vuelven $\theta \rightarrow \frac{1}{2}$ y $\theta > \frac{1}{2}$, respectivamente. Si un valor muestral es igual a μ_0, lo cual puede suceder cuando tratamos con datos redondeados aun cuando la población sea continua, simplemente lo descartamos.

Para realizar una prueba del signo cuando el tamaño de la muestra es muy pequeño, nos referimos directamente a una tabla de probabilidades binomiales como la tabla I; cuando el tamaño de la muestra es grande, usamos la aproximación normal de la distribución binomial.

EJEMPLO 16.1

Éstas son las mediciones de la resistencia a la ruptura, en libras, de cierta clase de cinta de algodón de 2 pulgadas:

163	165	160	189	161	171	158	151	169	162
163	139	172	165	148	166	172	163	187	173

Use la prueba del signo para probar la hipótesis nula $\mu = 160$ contra la hipótesis alternativa $\mu > 160$ en el nivel 0.05 de significancia.

Solución

1. H_0: $\mu = 160$
   H_1: $\mu > 160$
   $\alpha = 0.05$

2'. Use la estadística de prueba X, el número observado de signos positivos.

3'. Reemplace cada valor que excede a 160 con un signo más, cada valor menor de 160 con un signo menos, y descarte el valor único que es igual a 160, obtenemos

$$+ \; + \; + \; + \; + \; - \; - \; + \; + \; + \; - \; + \; + \; - \; + \; + \; + \; + \; +$$

de manera que $n = 19$ y $x = 15$. A partir de la tabla I encontramos que $P(X \geq 15) = 0.0095$ para $\theta = \frac{1}{2}$.

4'. Puesto que el valor P, 0.0095, es menor que 0.05, se debe rechazar la hipótesis nula, y concluimos que la media de la resistencia a la ruptura de la clase dada de cinta excede a 160 libras. ▲

EJEMPLO 16.2

Los datos siguientes, en toneladas, son las cantidades de óxidos de azufre emitidos por una planta industrial grande en 40 días:

17	15	20	29	19	18	22	25	27	9
24	20	17	6	24	14	15	23	24	26
19	23	28	19	16	22	24	17	20	13
19	10	23	18	31	13	20	17	24	14

Use la prueba del signo para probar la hipótesis nula $\mu = 21.5$ contra la hipótesis alternativa $\mu < 21.5$ en el nivel 0.01 de significancia.

Solución

1. H_0: $\mu = 21.5$
   H_1: $\mu < 21.5$
   $\alpha = 0.01$
2. Rechace la hipótesis nula si $z \leq -z_{0.01} = -2.33$, donde
$$z = \frac{x - n\theta}{\sqrt{n\theta(1-\theta)}}$$
   con $\theta = \frac{1}{2}$, y x es el número de signos más (valores que exceden a 21.5).
3. Puesto que $n = 40$ y $x = 16$, obtenemos $n\theta = 40 \cdot \frac{1}{2} = 20$, $\sqrt{n\theta(1-\theta)} = \sqrt{40(0.5)(0.5)} = 3.16$, y por tanto
$$z = \frac{16 - 20}{3.16} = -1.26$$
4. Puesto que $z = -1.26$ excede a -2.33, no se puede rechazar la hipótesis nula. ▲

También se puede usar la prueba del signo cuando tratamos con datos asociados en parejas, como en los ejercicios 13.31 y 13.32. En tales problemas, cada par de los valores de la muestra se reemplaza con un signo más si la diferencia entre las observaciones asociadas en parejas es positiva (esto es, si el primer valor excede al segundo valor) y por un signo menos si la diferencia entre las observaciones asociadas en parejas es negativa (esto es, si el primer valor es menor que el segundo valor), y se descarta si la diferencia es cero. Para probar la hipótesis nula de que dos poblaciones simétricas continuas tienen medias iguales (o que dos poblaciones continuas tienen medianas iguales), podemos así usar la prueba del signo, que en relación con esta clase de problemas se conoce como **la prueba del signo de muestras asociadas en parejas**. Cuando la prueba del signo se usa como en los ejemplos 16.1 y 16.2 nos referimos a ella como la **prueba del signo de una muestra**.

EJEMPLO 16.3

Para determinar la eficacia de un nuevo sistema de control de tránsito, se observó el número de accidentes que ocurrieron en 12 intersecciones peligrosas durante 4 semapara variables aleatorias y las letras minúsculas correspondientes para sus valores. Esto evita la confusión entre las estadísticas usadas aquí y la estadística t del capítulo 13.)

Puesto que la suma de T^+ y T^- es siempre $\frac{n(n+1)}{2}$ y ambas son valores de variables aleatorias que asumen valores en el intervalo de 0 a $\frac{n(n+1)}{2}$, con distribuciones que son simétricas alrededor de $\frac{n(n+1)}{4}$, podemos dibujar la relación entre las distribuciones de las variables aleatorias correspondientes a T^+, T^- y T como en la figura 16.1 para $n = 5$.

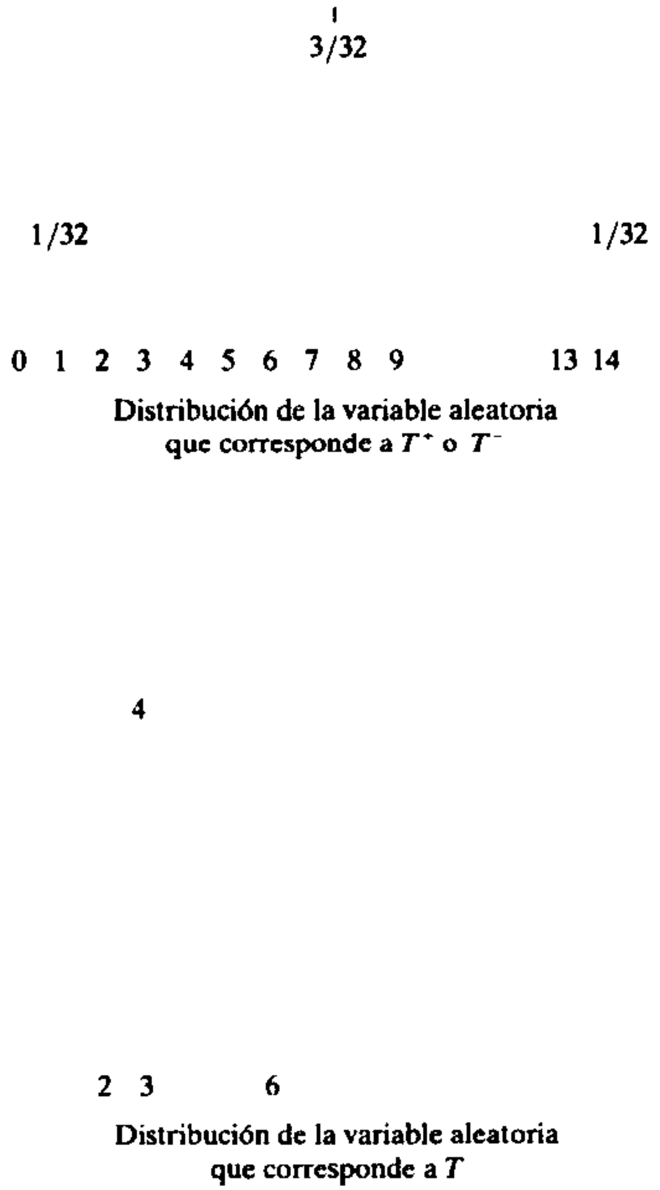

Figura 16.1 Distribuciones de variables aleatorias que corresponden a T^+, T^-, y T para $n = 5$.

Antes	*Después*
147.0	137.9
183.5	176.2
232.1	219.0
161.6	163.8
197.5	193.5
206.3	201.4
177.0	180.6
215.4	203.2
147.7	149.0
208.1	195.4
166.8	158.5
131.9	134.4
150.3	149.3
197.2	189.1
159.8	159.1
171.7	173.2

Use la prueba de rangos con signo para probar en el nivel 0.05 de significancia si el régimen alimenticio de reducción de peso es eficaz.

Solución

1. H_0: $\mu_1 = \mu_2$
   H_1: $\mu_1 > \mu_2$
   $\alpha = 0.05$
2. Rechace la hipótesis nula si $z \geqq z_{0.05} = 1.645$, donde

$$z = \frac{T^+ - \mu}{\sigma}$$

   y μ y σ^2 están dadas por las fórmulas del teorema 16.1.
3. Las diferencias entre los pares respectivos son 9.1, 7.3, 13.1, −2.2, 4.0, 4.9, −3.6, 12.2, −1.3, 12.7, 8.3, −2.5, 1.0, 8.1, 0.7 y −1.5, y si sus valores absolutos se ordenan, encontramos que las diferencias positivas ocupan los rangos 13, 10, 16, 8, 9, 14, 15, 12, 2, 11 y 1. Así,

$$T^+ = 13 + 10 + 16 + 8 + 9 + 14 + 15 + 12 + 2 + 11 + 1$$

$$= 111$$

   Puesto que $\mu = \frac{16 \cdot 17}{4} = 68$ y $\sigma^2 = \frac{16 \cdot 17 \cdot 33}{24} = 374$, obtenemos

$$z = \frac{111 - 68}{\sqrt{374}} = 2.22$$

4. Puesto que $z = 2.22$ excede a $z_{0.05} = 1.645$, se debe rechazar la hipótesis nula; concluimos que el régimen alimenticio es, en realidad, eficaz para reducir de peso. ▲

EJERCICIOS

16.1 Se toma una muestra aleatoria de tamaño $n = 2$ para probar si una población normal tiene la media $\mu = 0$.

(a) Si los valores observados de la muestra son x_1 y x_2 con $x_1 > x_2 > 0$, demuestre que la estadística para la prueba t de una muestra se puede escribir como

$$t = \frac{x_1 + x_2}{x_1 - x_2}$$

(b) Si el punto decimal se mueve erróneamente un lugar a la derecha cuando se registra x_1, encuentre una expresión para t', el valor correspondiente de la estadística t, y verifique que

$$1 < t' < t$$

16.2 Muestre que bajo la hipótesis nula de la sección 16.3, T^+ es un valor de una variable aleatoria cuya distribución es simétrica alrededor de $\frac{n(n+1)}{4}$.

16.3 Con respecto a la prueba de rangos con signo, encuentre la media y la varianza de la variable aleatoria cuyos valores están dados por $T^+ - T^-$.

16.4 Explique por qué, entre otros, está en blanco el elemento de la tabla **X** para $n = 5$ en la columna para $T_{0.02}$.

APLICACIONES

16.5 Éstas son las cantidades de tiempo, en minutos, que tardó a una muestra aleatoria de 20 técnicos en realizar cierta tarea: 18.1, 20.3, 18.3, 15.6, 22.5, 16.8, 17.6, 16.9, 18.2, 17.0, 19.3, 16.5, 19.5, 18.6, 20.0, 18.8, 19.1, 17.5, 18.5 y 18.0. Suponiendo que esta muestra vino de una población simétrica continua, use la prueba del signo en el nivel 0.05 de significancia para probar la hipótesis nula que la media de la población es 19.4 minutos contra la hipótesis alternativa que no es 19.4 minutos. Realice la prueba usando

(a) la tabla I;

(b) la aproximación normal a la distribución binomial.

16.6 Rehaga el ejercicio 16.5 usando la prueba de rangos con signo basada en la tabla **X**.

16.7 Éstas son las cantidades de dinero (en dólares) que gastaron 16 personas en un parque de diversiones: 20.15, 19.85, 23.75, 18.63, 21.09, 25.63, 16.65, 19.27, 18.80, 21.45, 20.29, 19.51, 23.80, 20.00, 17.48 y 19.11. Suponiendo que ésta es una muestra aleatoria de una población simétrica y que la probabilidad de que una persona gastará $19.00 exactamente es extremadamente pequeña, use la prueba del de significancia para probar la hipótesis nula $\mu = \mu_0$ contra la hipótesis alternativa

(a) $\mu \neq \mu_0$; (b) $\mu > \mu_0$; (c) $\mu < \mu_0$?

16.16 Rehaga el ejercicio 16.15 cambiando el nivel de significancia a 0.01.

16.17 En una muestra aleatoria tomada en un parque público, se tardaron 38, 43, 36, 29, 44, 28, 40, 50, 39, 47 y 33 minutos en jugar un partido de tenis. Use la prueba de rangos con signo en el nivel 0.05 de significancia para probar si en promedio se tardan 35 minutos en jugar un partido de tenis en ese parque público.

16.18 Una muestra de 24 maletas que transporta una línea aérea en vuelos transoceánicos pesó 32.0, 46.4, 48.1, 27.7, 35.5, 52.6, 66.0, 41.3, 49.9, 36.1, 50.0, 44.7, 48.2, 36.9, 40.8, 35.1, 63.3, 42.5, 52.4, 40.9, 38.6, 43.2, 41.7 y 35.6 libras. Pruebe en el nivel 0.05 de significancia si la media del peso de las maletas que transportó la línea aérea en esos vuelos es 37.0 libras usando la prueba de rangos con signo basada en

(a) la tabla X; (b) los resultados del teorema 16.1.

16.19 Lo siguiente es una muestra aleatoria de los IQ de maridos y esposas: 108 y 103, 104 y 116, 103 y 106, 112 y 104, 99 y 99, 105 y 94, 102 y 110, 112 y 128, 119 y 106, 106 y 103, 125 y 120, 96 y 98, 107 y 117, 115 y 130, 101 y 100, 110 y 101, 103 y 96, 105 y 99, 124 y 120, y 113 y 116. Pruebe en el nivel 0.05 de significancia si los maridos y las mujeres son en promedio igualmente inteligentes en la población muestreada usando la prueba de rangos con signo basada en

(a) la tabla X; (b) los resultados del teorema 16.1.

16.4 PRUEBAS DE SUMA DE RANGOS: LA PRUEBA *U*

En esta sección presentaremos una alternativa no paramétrica a la prueba t de dos muestras, que se llama la **prueba U**, la **prueba Wilcoxon**, o la **prueba Mann-Whitney**, nombradas en honor a los estadísticos que contribuyeron a su desarrollo. Sin tener que suponer que las dos poblaciones muestreadas tienen distribuciones normales, podremos probar la hipótesis nula de que estamos muestreando poblaciones continuas idénticas contra la alternativa de que las dos poblaciones tienen medias desiguales.

Para ilustrar el procedimiento, suponga que queremos comparar dos clases de señales luminosas de emergencia con base en los siguientes tiempos de iluminación (redondeadas al décimo de minuto más cercano):

Marca A : 14.9, 11.3, 13.2, 16.6, 17.0, 14.1, 15.4, 13.0, 16.9
Marca B : 15.2, 19.8, 14.7, 18.3, 16.2, 21.2, 18.9, 12.2, 15.3, 19.4

Al arreglar estos valores en forma conjunta (como si fueran una muestra) en un orden creciente en magnitud y asignarles en este orden los rangos 1, 2, 3, ... y 19, encontramos que los valores de la primera muestra (marca A) ocupan los rangos 1, 3, 4, 5, 7, 10, 12, 13 y 14, en tanto que los de la segunda muestra (marca B) ocupan los rangos, 2, 6, 8, 9, 11, 15, 16, 17, 18 y 19. Si hubiese habido empates, les habríamos asignado a cada una de las observaciones empatadas la media de los rangos que hubiesen ocupado conjuntamente.

Si hay una diferencia notable entre las medias de las dos poblaciones, es probable que la mayoría de los rangos inferiores vaya a los valores de una muestra, en tan-y la varianza

$$\sigma^2 = \frac{n_1 n_2(n_1 + n_2 + 1)}{12}$$

Demostración. Bajo la hipótesis nula de que las dos muestras vienen de poblaciones idénticas que son continuas (de manera que la probabilidad de que habrá cualesquier empates sea cero), W_1 es la suma de n_1 enteros positivos seleccionados al azar entre los primeros $n_1 + n_2$ enteros positivos. Al hacer uso de los resultados del inciso (c) del ejercicio 8.13 con $n = n_1$ y $N = n_1 + n_2$, encontramos que W_1 es el valor de una variable aleatoria con la media

$$\frac{n_1(n_1 + n_2 + 1)}{2}$$

y la varianza

$$\frac{n_1 n_2(n_1 + n_2 + 1)}{12}$$

Puesto que $U_1 = W_1 - \frac{n_1(n_1 + 1)}{2}$, se sigue que la media y la varianza de variable aleatoria que corresponden a U_1 son

$$\mu = \frac{n_1(n_1 + n_2 + 1)}{2} - \frac{n_1(n_1 + 1)}{2} = \frac{n_1 n_2}{2}$$

y

$$\sigma^2 = \frac{n_1 n_2(n_1 + n_2 + 1)}{12}$$

También, puesto que $U_1 + U_2$ siempre es igual a $n_1 n_2$, la media y la varianza de la variable aleatoria que corresponde a U_2 son iguales a las de la variable aleatoria que corresponde a U_1 [véase el inciso (a) del ejercicio 16.20]. ▼

EJEMPLO 16.7

A continuación se muestran los aumentos de peso (en libras) de dos muestras aleatorias de pavos jóvenes alimentados con dos dietas diferentes pero, en otros aspectos, mantenidos bajo condiciones idénticas:

Dieta 1: 16.3, 10.1, 10.7, 13.5, 14.9, 11.8, 14.3, 10.2, 12.0, 14.7, 23.6, 15.1, 14.5, 18.4, 13.2, 14.0

Dieta 2: 21.3, 23.8, 15.4, 19.6, 12.0, 13.9, 18.8, 19.2, 15.3, 20.1, 14.8, 18.9, 20.7, 21.1, 15.8, 16.2

Use la prueba U en el nivel 0.01 de significancia para probar la hipótesis nula que las dos poblaciones muestreadas son idénticas contra la hipótesis alternativa que en promedio la segunda dieta produce un mayor aumento de peso.

$$H = \frac{12}{18 \cdot 19}\left(\frac{84^2}{6} + \frac{55.5^2}{7} + \frac{31.5^2}{5}\right) - 3 \cdot 19$$
$$= 6.67$$

4. Puesto que $H = 6.67$ excede a $\chi^2_{0.05,2} = 5.991$, se debe rechazar la hipótesis nula; concluimos que los tres métodos no son igualmente eficaces. ▲

EJERCICIOS

16.20 Muestre que

(a) $U_1 + U_2 = n_1 n_2$ para cualquier par de valores de las variables aleatorias;

(b) ambas variables aleatorias correspondientes a U_1 y U_2 asumen valores en el rango de 0 a $n_1 n_2$.

16.21 Demuestre que la distribución de la variable aleatoria que corresponde a W_1 es simétrica alrededor de

$$\frac{n_1(n_1 + n_2 + 1)}{2}$$

y por tanto que la distribución de la variable aleatoria que corresponde a U_1 es simétrica alrededor de $\frac{n_1 n_2}{2}$. (*Sugerencia*: Ordene los datos combinados tanto en orden creciente como decreciente de magnitud.)

16.22 Verifique que U_1 y U_2 también están dadas por

$$U_1 = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - W_2$$

y

$$U_2 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - W_1$$

16.23 Si $X_1, X_2, \ldots, X_{n_1}$ y $Y_1, Y_2, \ldots, Y_{n_2}$ son variables aleatorias independientes, podemos probar la hipótesis nula de que vienen de poblaciones continuas idénticas con base en la estadística U de Mann-Whitney, la cual es simplemente el número de pares (x_i, y_j) para las cuales $x_i > y_j$. En forma simbólica,

$$U = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} d_{ij}$$

donde

$$d_{ij} = \begin{cases} 1 & \text{si } x_i > y_j \\ 0 & \text{si } x_i < y_j \end{cases}$$

para $i = 1, 2, \ldots, n_1$ y $j = 1, 2, \ldots, n_2$. Muestre que esta estadística U de Mann-Whitney es la misma que la estadística U_1 de la sección 16.4.

16.24 Verifique que la estadística Kruskal-Wallis en la página 544 es equivalente a

$$H = \frac{12}{n(n+1)} \cdot \sum_{i=1}^{k} n_i \left(\frac{R_i}{n_i} - \frac{n+1}{2} \right)^2$$

16.25 Muestre que si un análisis de la varianza en un sentido se realiza sobre los rangos de las observaciones en vez de las observaciones mismas, se vuelve equivalente a la prueba basada en la estadística H.

APLICACIONES

16.26 Éstas son cifras acerca del número de robos cometidos en una ciudad en muestras aleatorias de seis días en la primavera y seis días en el otoño:

Primavera:	36, 25, 32, 38, 28, 35
Otoño:	27, 20, 15, 29, 18, 22

Use la prueba U en el nivel 0.05 de significancia para probar la afirmación que en promedio hay igualmente tantos robos por día en la primavera como en el otoño, contra la alternativa de que hay menos en el otoño.

16.27 Los siguientes son los números de dureza Rockwell obtenidos para seis fundiciones a troquel de aluminio seleccionadas aleatoriamente del lote de producción A y ocho del lote de producción B:

Lote de producción A:	75, 56, 63, 70, 58, 74
Lote de producción B:	63, 85, 77, 80, 86, 76, 72, 82

Use la prueba U para probar en el nivel 0.05 de significancia para probar si las fundiciones del lote de producción B son en promedio igualmente duras o si son más duras que las del lote de producción A.

16.28 A continuación se muestra el número de minutos que a una muestra aleatoria de 15 hombres y 12 mujeres tardó en terminar una prueba escrita dada para la renovación de sus licencias de manejar:

Hombres:	9.9, 7.4, 8.9, 9.1, 7.7, 9.7, 11.8, 9.2, 10.0, 10.2, 9.5, 10.8, 8.0, 11.0, 7.5
Mujeres:	8.6, 10.9, 9.8, 10.7, 9.4, 10.3, 7.3, 11.5, 7.6, 9.3, 8.8, 9.6

Use la prueba U basada en la tabla XI en el nivel 0.05 de significancia para decidir si aceptamos la hipótesis nula $\mu_1 = \mu_2$ o la hipótesis alternativa $\mu_1 \neq \mu_2$, donde μ_1 y μ_2 son los promedios de la cantidad de tiempo que tardan los hombres y las mujeres en terminar la prueba.

16.29 Rehaga el ejercicio 16.28 usando la aproximación normal a la distribución de la estadística de prueba.

16.30 Un examen diseñado para medir el conocimiento básico sobre la historia de Estados Unidos se aplicó a muestras aleatorias de alumnos de primer año de dos universidades importantes, y sus calificaciones fueron:

dosis de 0.5 *mg*:	8.2	10.0	10.2	13.7	14.0	7.8
	12.7	10.9				
dosis de 1.0 *mg*:	9.7	13.1	11.0	7.5	13.3	12.5
	8.8	12.9	7.9	10.5		
dosis de 1.5 *mg*:	12.0	7.2	8.0	9.4	11.3	9.0
	11.5	8.5				

Use la prueba H en el nivel 0.01 de significancia para probar la hipótesis nula de que las diferencias en dosis no tienen efecto en el periodo que tardan los conejillos de indias en dormirse.

16.6 PRUEBAS BASADAS EN CORRIDAS

Hay varios métodos no paramétricos para probar la aleatoriedad de los datos observados con base en el orden en que se obtuvieron. La técnica que describiremos aquí se basa en la **teoría de las corridas**, donde una **corrida** es una sucesión de letras idénticas (u otra clase de símbolos) precedida o seguida por diferentes letras o por ninguna letra. Para ilustrarlo, considere los siguientes arreglos de piezas defectuosas, d, y no defectuosas, n, que cierta máquina produjo en el orden dado:

$$\underbrace{n\ n\ n\ n\ n}\ \underbrace{d\ d\ d\ d}\ \underbrace{n\ n\ n\ n\ n\ n\ n\ n\ n\ n}\ \underbrace{d\ d}\ \underbrace{n\ n}\ \underbrace{d\ d\ d\ d}\ \underbrace{n}\ \underbrace{d\ d}\ \underbrace{n\ n}$$

Al usar corchetes para combinar las letras que constituyen una corrida, encontramos que hay primero una corrida de cinco n, después una corrida de cuatro d, después una corrida de diez n, ... y finalmente una corrida de dos n; en total, hay nueve corridas de longitud variable.

El número total de corridas que aparece en un arreglo de esta clase es, a menudo, una buena indicación de una posible falta de aleatoriedad. Si hay muy pocas corridas, podríamos sospechar una agrupación o apiñamiento, o quizá una tendencia; si hay demasiadas corridas, podríamos sospechar algún tipo de patrón alternante que se repite. En nuestra ilustración, parece haber una agrupación definitiva, las piezas defectuosas parecen venir en grupos; pero queda por ver si esto es significativo o si se puede atribuir al azar.

Para encontrar la probabilidad de que n_1 letras de una clase y n_2 letras de otra clase formarán u corridas cuando cada uno de los $\binom{n_1 + n_2}{n_1}$ arreglos posibles de estas letras se considera como igualmente probable, investiguemos primero el caso donde u es par, esto es, cuando $u = 2k$ y k es un entero positivo. En este caso tendrá que haber k corridas de cada clase alternando entre sí. Para encontrar el número de formas en que n_1 letras pueden formar k corridas, consideremos primero el caso muy simple donde tenemos cinco letras c que se van a dividir en tres corridas. Al usar barras verticales para separar las cinco letras en tres corridas, encontramos que hay *seis* posibilidades

$$c|c|ccc \quad c|cc|cc \quad c|ccc|c$$

$$cc|c|cc \quad cc|cc|c \quad ccc|c|c$$

que corresponden a las $\binom{4}{2}$ formas en que podemos colocar dos barras verticales en dos de los cuatro espacios entre las cinco c. Por la misma razón, hay $\binom{n_1 - 1}{k - 1}$ formas en las que n_1 letras de la primera clase pueden formar k corridas y $\binom{n_2 - 1}{k - 1}$ formas en las que n_2 letras de la segunda clase pueden formar k corridas, y se sigue que hay en total $2\binom{n_1 - 1}{k - 1}\binom{n_2 - 1}{k - 1}$ formas en las que estas $n_1 + n_2$ letras pueden formar $2k$ corridas. El factor 2 se explica por el hecho que cuando combinamos las dos clases de corridas de manera que puedan alternar, podemos empezar ya sea con una corrida de la primera clase de letra o con una corrida de la segunda clase. Así, cuando $u = 2k$ (donde k es un entero positivo), la probabilidad de obtener todas esas corridas es

$$f(u) = \frac{2\binom{n_1 - 1}{k - 1}\binom{n_2 - 1}{k - 1}}{\binom{n_1 + n_2}{n_1}}$$

y se deja al lector para demostrar en el ejercicio 16.37 que argumentos similares nos llevan a

$$f(u) = \frac{\binom{n_1 - 1}{k}\binom{n_2 - 1}{k - 1} + \binom{n_1 - 1}{k - 1}\binom{n_2 - 1}{k}}{\binom{n_1 + n_2}{n_1}}$$

donde $u = 2k + 1$ (donde k es un entero positivo).

Cuando n_1 y n_2 son pequeños, las pruebas de aleatoriedad basadas en u suelen realizarse con las tablas especiales como la XII. Rechazamos la hipótesis nula de aleatoriedad en el nivel α de significancia si

$$u \leqq u'_{\alpha/2} \quad \text{o} \quad u \geqq u_{\alpha/2}$$

donde $u'_{\alpha/2}$ es el valor más grande para el cual la probabilidad de obtener un valor menor que, o igual a, éste no excede $\alpha/2$, y $u_{\alpha/2}$ es el valor más pequeño para el cual la probabilidad de obtener un valor mayor que, o igual a, éste no excede a $\alpha/2$.

EJEMPLO 16.9

Al revisar los olmos que se plantaron hace muchos años a lo largo de un camino vecinal, un funcionario del condado obtuvo los siguientes arreglos de árboles sanos, S, y enfermos, E:

$$S\ S\ S\ S\ E\ E\ E\ S\ S\ S\ S\ S\ S\ S\ E\ E\ S\ S\ E\ E\ E\ E$$

Pruebe en el nivel 0.05 de significancia si este arreglo se puede considerar como fortuito.

Solución

1. H_0: El arreglo es fortuito.
   H_1: El arreglo no es fortuito.
   $\alpha = 0.05$
2. Puesto que $n_1 = 13$ y $n_2 = 9$, rechace la hipótesis nula si $u \leqq 6$ o $u \geqq 17$, donde 6 y 17 son los valores correspondientes de $u'_{0.025}$ y $u_{0.025}$.
3. $u = 6$ por inspección de los datos.
4. Puesto que $u = 6$ es igual a $u'_{0.025} = 6$, se debe rechazar la hipótesis nula; el arreglo de olmos sanos y enfermos no es fortuito. Parece que los árboles enfermos vienen en grupos. ▲

Cuando n_1 y n_2 son ambos mayores que, o iguales a 10, se considera razonable suponer que la distribución de la variable aleatoria correspondiente a u se puede aproximar muy cercanamente con una curva normal. Para efectuar las corridas de prueba sobre la base de esta suposición, necesitamos los siguientes resultados.

TEOREMA 16.3 Bajo la hipótesis nula de aleatoriedad, la media y la varianza de la variable aleatoria que corresponde a u son

$$\mu = \frac{2n_1n_2}{n_1 + n_2} + 1$$

y

$$\sigma^2 = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}$$

Estos resultados se pueden obtener directamente con las fórmulas dadas en la página 549. Los detalles de esas pruebas, así como un enfoque alternativo que es más fácil, se pueden encontrar en el libro de J. D. Gibbons listado entre las referencias al final de este capítulo.

EJEMPLO 16.10

Lo siguiente es un arreglo de hombres, H, y mujeres, M, que hacen fila para comprar boletos de un concierto de "rock":

H M H M H H H M H M H H H M M H H H H M M H M H

H H M H H H M M M H M H H H M H M H H H H M M H

Pruebe por aleatoriedad en el nivel 0.05 de significancia.

Solución

1. H_0: El arreglo es fortuito.
   H_1: El arreglo no es fortuito.
   $\alpha = 0.05$

u es el número de corridas por arriba y por debajo de la mediana, y μ y σ^2 están dadas por las fórmulas del teorema 16.3.

3. Puesto que la mediana de las velocidades es 59.5, obtenemos el siguiente arreglo de a y b:

$$b\ b\ a\ b\ a\ a\ b\ b\ a\ b\ b\ b\ b\ a\ b\ a\ a\ a\ a\ a\ b\ b\ b\ b\ a$$

$$b\ b\ a\ b\ b\ b\ a\ a\ a\ b\ a\ a\ a\ b\ b\ b\ b\ b\ a\ a\ a\ a\ a\ a\ a$$

Entonces, puesto que $n_1 = 25$, $n_2 = 25$ y $u = 20$, obtenemos

$$\mu = \frac{2 \cdot 25 \cdot 25}{25 + 25} + 1 = 26$$

$$\sigma^2 = \frac{2 \cdot 25 \cdot 25(2 \cdot 25 \cdot 25 - 25 - 25)}{(25 + 25)^2(25 + 25 - 1)} = 12.2$$

y

$$z = \frac{20 - 26}{\sqrt{12.2}} = -1.72$$

4. Puesto que $z = -1.72$ cae entre -1.96 y 1.96, no se puede rechazar la hipótesis nula; no hay evidencia real para que la muestra no se pueda considerar como al azar. ▲

EJERCICIOS

16.37 Verifique la fórmula dada en la página 549 para la probabilidad de obtener u corridas cuando $u = 2k + 1$, donde k es un entero positivo.

16.38 Si una persona obtiene siete caras y tres cruces en 10 lanzamientos de una moneda balanceada, encuentre las probabilidades para 2, 3, 4, 5, 6 y 7 corridas.

16.39 Encuentre la probabilidad de que $n_1 = 6$ letras de una clase y $n_2 = 5$ letras de otra clase formarán al menos 8 corridas.

16.40 Si hay $n_1 = 8$ letras de una clase y $n_2 = 8$ letras de otra clase, ¿para cuántas corridas rechazaríamos la hipótesis nula de aleatoriedad en el nivel 0.01 de significancia?

APLICACIONES

16.41 Éste es el orden en que un corredor recibe órdenes de compra, C, y venta, V para cierta acción:

$$C\ C\ C\ C\ C\ C\ C\ C\ V\ V\ C\ V\ V\ V\ V\ V\ C\ C\ C\ C\ C$$

Pruebe por aleatoriedad en el nivel 0.05 de significancia.

16.42 Una conductora compra gasolina en una gasolinera Texaco, T, o en una gasolinera Mobil, M, y los siguientes arreglos muestran el orden de las gasolineras donde compró gasolina durante cierto periodo:

155, 146 y 158 durante un periodo de 33 años. Haga uso del hecho que la mediana es 138, pruebe en el nivel 0.05 de significancia si hay una tendencia verdadera.

16.51 Éstas son las ventas trimestrales durante seis años (en millones de dólares) de un fabricante de maquinaria pesada: 83.8, 102.5, 121.0, 90.5, 106.6, 104.8, 114.7, 93.6, 98.9, 96.9, 122.6, 85.6, 103.2, 96.9, 118.0, 92.1, 100.5, 92.9, 125.6, 79.2, 110.8, 95.1, 125.6 y 86.7. ¿En el nivel 0.05 de significancia, hay un verdadero patrón cíclico?

16.52 La teoría de las corridas también se puede usar como una alternativa de la prueba de suma de rangos de la sección 16.4, esto es, la prueba de la hipótesis nula que dos variables aleatorias independientes vienen de poblaciones continuas idénticas. Simplemente ordenados los datos conjuntamente, escribimos un 1 debajo de cada valor que pertenece a la primera muestra y un 2 debajo de cada valor que pertenece a la segunda muestra, y después probamos la aleatoriedad del arreglo resultante de 1 y 2. Si hay muy pocas corridas, esto bien puede explicarse por el hecho que las dos muestras vienen de poblaciones con medias desiguales. Con respecto a los datos en la página 539, use esta técnica para probar en el nivel 0.05 de significancia si las dos muestras vinieron de poblaciones continuas idénticas o si las dos poblaciones tienen medias desiguales.

16.7 EL COEFICIENTE DE CORRELACIÓN DE RANGOS

Puesto que las suposiciones que sustentan la prueba de significancia para los coeficientes de correlación de la sección 14.5 son más bien estrictas, a veces es preferible usar una alternativa no paramétrica. Muy popular entre esas medidas de asociación no paramétricas está el **coeficiente de correlación de rangos**, también llamado **coeficiente de correlación de rangos de Spearman**, r_S. Para un conjunto dado de datos asociados en parejas $\{(x_i, y_i);\ i = 1, 2, \ldots, n\}$, se obtiene al ordenar las x entre sí mismas y también las y, ambas de menor a mayor o de mayor a menor, y después sustituirlas en la siguiente fórmula.

DEFINICIÓN 16.1 El coeficiente de correlación de rangos está dado por

$$r_S = 1 - \frac{6 \cdot \sum_{i=1}^{n} d_i^2}{n(n^2 - 1)}$$

donde d_i es la diferencia entre los rangos asignados a x_i y y_i.

Cuando hay un empate en el rango, procedemos como antes y asignamos a las observaciones empatadas la media de los rangos que ocupan conjuntamente.

Cuando no hay empates en el rango, r_S es realmente igual al coeficiente de correlación r calculado para los rangos. Para verificar esto, sean r_i y s_i los rangos de x_i y y_i. Al hacer uso del hecho que la suma y la suma de los cuadrados de los primeros n enteros positivos son $\frac{n(n+1)}{2}$ y $\frac{n(n+1)(2n+1)}{6}$, respectivamente, encontramos que

$$\sum_{i=1}^{n} r_i = \sum_{i=1}^{n} s_i = \frac{n(n+1)}{2}$$

$$\sum_{i=1}^{n} r_i^2 = \sum_{i=1}^{n} s_i^2 = \frac{n(n+1)(2n+1)}{6}$$

$$\sum_{i=1}^{n} r_i s_i = \frac{n(n+1)(2n+1)}{6} - \frac{1}{2} \cdot \sum_{i=1}^{n} d_i^2$$

y si sustituimos estas expresiones en la fórmula para r, obtenemos la fórmula para r_S dada en la definición 16.1.

EJEMPLO 16.12

A continuación se da el número de horas que 10 estudiantes estudiaron para un examen y las puntuaciones que obtuvieron:

Número de horas estudiadas x	*Puntuación* y
8	56
5	44
11	79
13	72
10	70
5	54
18	94
15	85
2	33
8	65

Calcule r_S.

Solución

Al ordenar las x y las y y al proceder como en la siguiente tabla, obtenemos

Rango de x	*Rango de y*	*d*	d^2
6.5	7	−0.5	0.25
8.5	9	−0.5	0.25
4	3	1.0	1.00
3	4	1.0	1.00
5	5	0.0	0.00
8.5	8	0.5	0.25
1	1	0.0	0.00
2	2	0.0	0.00
10	10	0.0	0.00
6.5	6	0.5	0.25
			3.00

Entonces, la sustitución en la fórmula para r_S nos da

$$r_S = 1 - \frac{6 \cdot 3}{10(10^2 - 1)} = 0.98 \quad \blacktriangle$$

Como se puede ver a partir de este ejemplo, r_S se determina con facilidad; ciertamente, a veces se usa en vez de r principalmente a causa de su facilidad de cálculo. Si hubiésemos calculado r para los datos del ejemplo anterior, habríamos obtenido $r =$ 0.96, y esto está muy cercano a $r_S = 0.98$.

Para valores pequeños de $n(n \leqq 10)$, la prueba de la hipótesis nula de no correlación, ciertamente, la hipótesis nula de que las x y las y están asociadas aleatoriamente en parejas, se puede basar en las tablas especiales determinadas de las distribuciones muestrales exactas de R_S (véanse las referencias en la página 559). La mayoría de las veces, sin embargo, usamos el hecho que la distribución de muestreo de R_S puede aproximar muy cercanamente con la distribución normal, y para este fin necesitamos los siguientes resultados.

TEOREMA 16.4 Bajo la hipótesis nula de no correlación, la media y la varianza de R_S son

$$E(R_S) = 0 \quad \text{y} \quad \text{var}(R_S) = \frac{1}{n-1}$$

Una prueba de este teorema se puede encontrar en el libro de Gibbons que se encuentra entre las referencias al final de este capítulo. En el sentido estricto, el teorema se aplica cuando no hay empates, pero se puede usar a menos que el número de empates sea grande.

EJEMPLO 16.13

Con respecto al ejemplo 16.12, pruebe en el nivel 0.01 de significancia si el valor obtenido para r_S, 0.98, es significativo.

Solución

1. H_0: No hay correlación.
   H_1: Hay correlación.
   $\alpha = 0.01$
2. Rechace la hipótesis nula si $z \leqq -2.575$ o $z \geqq 2.575$, donde
$$z = r_S\sqrt{n-1}$$
3. Al sustituir $n = 10$ y $r_S = 0.98$, obtenemos
$$z = 0.98\sqrt{10-1} = 2.94$$
4. Puesto que $z = 2.94$ excede a 2.575, se debe rechazar la hipótesis nula; concluimos que hay una relación verdadera (positiva) entre el tiempo de estudio y las puntuaciones. ▲

EJERCICIOS

16.53 Dado un conjunto de $k$adas $(x_{11}, x_{12}, \ldots, x_{1k})$, $(x_{21}, x_{22}, \ldots, x_{2k})$, ... y $(x_{n1}, x_{n2}, \ldots, x_{nk})$, el alcance de su asociación, o concordancia, se puede medir por medio del **coeficiente de concordancia**:
$$W = \frac{12}{k^2n(n^2-1)} \cdot \sum_{i=1}^{n}\left[R_i - \frac{k(n+1)}{2}\right]^2$$
donde R_i es la suma de los rangos asignados a $x_{i1}, x_{i2}, \ldots$ y x_{ik} cuando las x con el segundo subíndice 1 se ordenan entre sí mismas y lo mismo se hace con las x con el segundo subíndice 2, ... y las x con el segundo subíndice k. ¿Cuáles son los valores máximos y mínimos de W, y qué reflejan con respecto a la concordancia, o falta de ella, de los valores de las k variables aleatorias?

APLICACIONES

16.54 Calcule r_S para los datos siguientes que representan las calificaciones de estadística, x, y las calificaciones de psicología, y, de 18 estudiantes

x	y	x	y
78	80	97	90
86	74	74	85
49	63	53	71
94	85	58	67
53	55	62	64
89	86	74	69
94	90	74	71
71	84	70	67
70	71	74	71

APÉNDICE

A

Sumas y productos

A.1 REGLAS PARA SUMAS Y PRODUCTOS
A.2 SUMAS ESPECIALES

A.1 REGLAS PARA SUMAS Y PRODUCTOS

Para simplificar las expresiones que incluyen sumas y productos, en estadística se usan ampliamente las notaciones $\sum$ y $\prod$. En la notación usual escribimos

$$\sum_{i=a}^{b} x_i = x_a + x_{a+1} + x_{a+2} + \cdots + x_b$$

y

$$\prod_{i=a}^{b} x_i = x_a \cdot x_{a+1} \cdot x_{a+2} \cdot \ldots \cdot x_b$$

para cualesquier enteros no negativos a y b con $a \leqq b$.

Cuando se trabaja con sumas o productos, a menudo es útil aplicar las siguientes reglas, las cuales se pueden verificar al escribir las expresiones respectivas en su totalidad, esto es, sin la notación $\sum$ o $\prod$:

TEOREMA A.1

1. $\displaystyle\sum_{i=1}^{n} kx_i = k \cdot \sum_{i=1}^{n} x_i$
2. $\displaystyle\sum_{i=1}^{n} k = nk$
3. $\displaystyle\sum_{i=1}^{n} (x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i$
4. $\displaystyle\prod_{i=1}^{n} kx_i = k^n \cdot \prod_{i=1}^{n} x_i$
5. $\displaystyle\prod_{i=1}^{n} k = k^n$

6. $\prod_{i=1}^{n} x_i y_i = \left(\prod_{i=1}^{n} x_i\right)\left(\prod_{i=1}^{n} y_i\right)$

7. $\ln \prod_{i=1}^{n} x_i = \sum_{i=1}^{n} \ln x_i$

En la estadística también se usan ampliamente sumas dobles, sumas triples, ..., y si aplicamos repetidamente la definición de $\sum$ dada arriba, tenemos, por ejemplo

$$\begin{aligned}\sum_{i=1}^{m} \sum_{j=1}^{n} x_{ij} &= \sum_{i=1}^{m} (x_{i1} + x_{i2} + \cdots + x_{in}) \\ &= (x_{11} + x_{12} + \cdots + x_{1n}) \\ &\quad + (x_{21} + x_{22} + \cdots + x_{2n}) \\ &\quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ &\quad + (x_{m1} + x_{m2} + \cdots + x_{mn})\end{aligned}$$

Advierta que cuando las x_{ij} se acomodan así en un arreglo rectangular, el primer subíndice denota el renglón al cual pertenece el elemento en particular, y el segundo subíndice denota la columna.

Cuando trabajamos con sumas dobles, el siguiente teorema es de especial interés; es consecuencia inmediata de la expansión polinomial de

$$(x_1 + x_2 + \cdots + x_n)^2$$

TEOREMA A.2

$$\sum_{i<j}\sum x_i x_j = \frac{1}{2}\left[\left(\sum_{i=1}^{n} x_i\right)^2 - \sum_{i=1}^{n} x_i^2\right]$$

donde

$$\sum_{i<j}\sum x_i x_j = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} x_i x_j$$

A.2 SUMAS ESPECIALES

En la teoría de la estadística no paramétrica, particularmente cuando tratamos con sumas de rangos, frecuentemente necesitamos expresiones para sumas de potencias de los primeros n enteros positivos, esto es, expresiones para

$$S(n, r) = 1^r + 2^r + 3^r + \cdots + n^r$$

A.3 Dado $x_1 = 1$, $x_2 = 3$, $x_3 = -2$, $x_4 = 4$, $x_5 = -1$, $x_6 = 2$, $x_7 = 1$ y $x_8 = 2$, encuentre

(a) $\sum_{i=1}^{8} x_i$; (b) $\sum_{i=1}^{8} x_i^2$.

A.4 Dado $x_1 = 3$, $x_2 = 4$, $x_3 = 5$, $x_4 = 6$, $x_5 = 7$, $f_1 = 3$, $f_2 = 7$, $f_3 = 10$, $f_4 = 5$ y $f_5 = 2$, encuentre

(a) $\sum_{i=1}^{5} x_i$; (b) $\sum_{i=1}^{5} f_i$;

(c) $\sum_{i=1}^{5} x_i f_i$; (d) $\sum_{i=1}^{5} x_i^2 f_i$.

A.5 Dado $x_1 = 2$, $x_2 = -3$, $x_3 = 4$, $x_4 = -2$, $y_1 = 5$, $y_2 = -3$, $y_3 = 2$ y $y_4 = -1$, encuentre

(a) $\sum_{i=1}^{4} x_i$; (b) $\sum_{i=1}^{4} y_i$;

(c) $\sum_{i=1}^{4} x_i^2$; (d) $\sum_{i=1}^{4} y_i^2$; (e) $\sum_{i=1}^{4} x_i y_i$.

A.6 Dado $x_{11} = 3$, $x_{12} = 1$, $x_{13} = -2$, $x_{14} = 2$, $x_{21} = 1$, $x_{22} = 4$, $x_{23} = -2$, $x_{24} = 5$, $x_{31} = 3$, $x_{32} = -1$, $x_{33} = 2$ y $x_{34} = 3$, encuentre

(a) $\sum_{i=1}^{3} x_{ij}$ separadamente para $j = 1, 2, 3$, y 4;

(b) $\sum_{j=1}^{4} x_{ij}$ separadamente para $i = 1, 2$, y 3.

A.7 Con referencia al ejercicio A.6, evalúe la suma doble $\sum_{i=1}^{3} \sum_{j=1}^{4} x_{ij}$ usando

(a) los resultados del inciso (a) de ese ejercicio;

(b) los resultados del inciso (b) de ese ejercicio.

APÉNDICE

B

Distribuciones de probabilidad especiales

B.1 DISTRIBUCIÓN DE BERNOULLI

$$f(x; \theta) = \theta^x(1-\theta)^{1-x} \qquad \text{para } x = 0, 1$$

Parámetro: $0 < \theta < 1$

Media y varianza: $\mu = \theta$ y $\sigma^2 = \theta(1-\theta)$

B.2 DISTRIBUCIÓN BINOMIAL

$$b(x; n, \theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x} \qquad \text{para } x = 0, 1, 2, \ldots, n$$

Parámetros: n es un entero positivo y $0 < \theta < 1$

Media y varianza: $\mu = n\theta$ y $\sigma^2 = n\theta(1-\theta)$

B.3 DISTRIBUCIÓN UNIFORME DISCRETA (CASO ESPECIAL)

$$f(x; k) = \frac{1}{k} \qquad \text{para } x = 1, 2, \ldots, k$$

Parámetro: k es un entero positivo

Media y varianza: $\mu = \frac{k+1}{2}$ y $\sigma^2 = \frac{k^2-1}{12}$

APÉNDICE

C

Densidades de probabilidad especiales

C.1 DISTRIBUCIÓN BETA

$$f(x; \alpha, \beta) = \begin{cases} \dfrac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \cdot \Gamma(\beta)} x^{\alpha-1}(1 - x)^{\beta-1} & \text{para } 0 < x < 1 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

Parámetros: $\alpha > 0$ y $\beta > 0$

Media y varianza: $\mu = \dfrac{\alpha}{\alpha + \beta}$ y $\sigma^2 = \dfrac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$

C.2 DISTRIBUCIÓN DE CAUCHY

$$p(x; \alpha, \beta) = \frac{\frac{\beta}{\pi}}{(x - \alpha)^2 + \beta^2}$$

Parámetros: $-\infty < \alpha < \infty$ y $\beta > 0$

Media y varianza: No existen

C.3 DISTRIBUCIÓN JI CUADRADA

$$f(x; \nu) = \begin{cases} \dfrac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{\frac{\nu-2}{2}} e^{-\frac{x}{2}} & \text{para } x > 0 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

Parámetros: ν es un entero positivo

Media y varianza: $\mu = \nu$ y $\sigma^2 = 2\nu$

C.4 DISTRIBUCIÓN EXPONENCIAL

$$g(x;\theta) = \begin{cases} \dfrac{1}{\theta}e^{-x/\theta} & \text{para } x > 0 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

Parámetro: $\theta > 0$

Media y varianza: $\mu = \theta$ y $\sigma^2 = \theta^2$

C.5 DISTRIBUCIÓN F

$$g(f) = \begin{cases} \dfrac{\Gamma\left(\dfrac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\dfrac{\nu_1}{2}\right)\Gamma\left(\dfrac{\nu_2}{2}\right)}\left(\dfrac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} \cdot f^{\frac{\nu_1}{2}-1}\left(1+\dfrac{\nu_1}{\nu_2}f\right)^{-\frac{1}{2}(\nu_1+\nu_2)} & \text{para } f > 0 \\ = 0 & \text{en cualquier otra parte} \end{cases}$$

Parámetros: $\nu_1 > 0$ y $\nu_2 > 0$

Media: $\mu = \dfrac{\nu_2}{\nu_2 - 2}$

C.6 DISTRIBUCIÓN GAMMA

$$f(x) = \begin{cases} \dfrac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1}e^{-x/\beta} & \text{para } x > 0 \\ 0 & \text{en cualquier otra parte} \end{cases}$$

Parámetros: $\alpha > 0$ y $\beta > 0$

Media y varianza: $\mu = \alpha\beta$ y $\sigma^2 = \alpha\beta^2$

C.7 DISTRIBUCIÓN NORMAL

$$m(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad \text{para } -\infty < x < \infty$$

Parámetros: μ y $\sigma > 0$

Media y varianza: $\mu = \mu$ y $\sigma^2 = \sigma^2$

Tablas estadísticas

		θ						
		.15	.20	.25	.30			
8		.0026	.0092	.0231	.0467		.1719	.2188
		.0002	.0011	.0038	.0100		.0703	.1094
		.0000	.0001	.0004	.0012		.0164	.0312
		.0000	.0000	.0000	.0001		.0017	.0039
9	0	.2316	.1342	.0751	.0404		.0046	.0020
	1	.3679	.3020	.2253	.1556		.0339	.0176
	2	.2597	.3020	.3003	.2668		.1110	.0703
	3	.1069	.1762	.2336	.2668		.2119	.1641
	4	.0283	.0061	.1168	.1715		.2600	.2461
	5	.0050	.0165	.0389	.0735		.2128	.2461
	6	.0006	.0028	.0087	.0210		.1160	.1641
	7	.0000	.0003	.0012	.0039		.0407	.0703
	8	.0000	.0000	.0001	.0004		.0083	.0176
	9	.0000	.0000	.0000	.0000		.0008	.0020
10	0	.1969	.1074	.0563	.0282		.0025	.0010
	1	.3474	.2684	.1877	.1211		.0207	.0098
	2	.2759	.3020	.2816	.2335		.0763	.0439
	3	.1298	.2013	.2503	.2668		.1665	.1172
	4	.0401	.0881	.1460	.2001		.2384	.2051
	5	.0085	.0264	.0584	.1029		.2340	.2461
	6	.0012	.0055	.0162	.0368		.1596	.2051
	7	.0001	[illegible]	[illegible]	.0090		.0746	.1172
	8	.0000	.0001	.0004	.0014		.0229	.0439
	9	.0000	.0000	.0000	.0001		.0042	.0098
	10	.0000	.0000	.0000	.0000		.0003	.0016
11	0	.1673	.0859	.0422	.0198		.0014	.0005
	1	.3248	.2362	.1549	.0932		.0125	.0054
	2	.2866	.2953	.2581	.1998		.0513	.0269
	3	.1517	.2215	.2581	.2568		.1259	.0806
	4	.0536	.1107	.1721	.2201		.2060	.1611
	5	.0132	.0388	.0803	.1321		.2360	.2256
	6	.0023	.0097	.0268	.0566		.1931	.2256
	7	.0003	.0017	.0064	.0173		.1128	.1611
	8	.0000	.0002	.0011	.0037		.0462	.0806
	9	.0000	.0000	.0001	.0005	.0018	.0126	.0269
		.0000	.0000	.0000	.0000	.0002		.0054
		.0000	.0000	.0000	.0000	.0000		.0005
12								.0002
								.0029
								.0161
								.0537
								.1208
								.1934
								.2256
								.1934
								.1208
								.0537

TABLA I (continuación)

		θ									
n	x	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50
18	10	.0000	.0000	.0001	.0008	.0042	.0149	.0385	.0771	.1248	.1669
	11	.0000	.0000	.0000	.0001	.0010	.0046	.0151	.0374	.0742	.1214
	12	.0000	.0000	.0000	.0000	.0002	.0012	.0047	.0145	.0354	.0708
	13	.0000	.0000	.0000	.0000	.0000	.0002	.0012	.0045	.0134	.0327
	14	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0011	.0039	.0117
	15	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0009	.0031
	16	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0006
	17	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001
	18	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
19	0	.3774	.1351	.0456	.0144	.0042	.0011	.0003	.0001	.0000	.0000
	1	.3774	.2852	.1529	.0685	.0268	.0093	.0029	.0008	.0002	.0000
	2	.1787	.2852	.2428	.1540	.0803	.0358	.0138	.0046	.0013	.0003
	3	.0533	.1796	.2428	.2182	.1517	.0869	.0422	.0175	.0062	.0018
	4	.0112	.0798	.1714	.2182	.2023	.1491	.0909	.0467	.0203	.0074
	5	.0018	.0266	.0907	.1636	.2023	.1916	.1468	.0933	.0497	.0222
	6	.0002	.0069	.0374	.0955	.1574	.1916	.1844	.1451	.0949	.0518
	7	.0000	.0014	.0122	.0443	.0974	.1525	.1844	.1797	.1443	.0961
	8	.0000	.0002	.0032	.0166	.0487	.0981	.1489	.1797	.1771	.1442
	9	.0000	.0000	.0007	.0051	.0198	.0514	.0980	.1464	.1771	.1762
	10	.0000	.0000	.0001	.0013	.0066	.0220	.0528	.0976	.1449	.1762
	11	.0000	.0000	.0000	.0003	.0018	.0077	.0233	.0532	.0970	.1442
	12	.0000	.0000	.0000	.0000	.0004	.0022	.0083	.0237	.0529	.0961
	13	.0000	.0000	.0000	.0000	.0001	.0005	.0024	.0085	.0233	.0518
	14	.0000	.0000	.0000	.0000	.0000	.0001	.0006	.0024	.0082	.0222
	15	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0005	.0022	.0074
	16	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0005	.0018
	17	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003
	18	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	19	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
20	0	.3585	.1216	.0388	.0115	.0032	.0008	.0002	.0000	.0000	.0000
	1	.3774	.2702	.1368	.0576	.0211	.0068	.0020	.0005	.0001	.0000
	2	.1887	.2852	.2293	.1369	.0669	.0278	.0100	.0031	.0008	.0002
	3	.0596	.1901	.2428	.2054	.1339	.0716	.0323	.0123	.0040	.0011
	4	.0133	.0898	.1821	.2182	.1897	.1304	.0738	.0350	.0139	.0046
	5	.0022	.0319	.1028	.1746	.2023	.1789	.1272	.0746	.0365	.0148
	6	.0003	.0089	.0454	.1091	.1686	.1916	.1712	.1244	.0746	.0370
	7	.0000	.0020	.0160	.0545	.1124	.1643	.1844	.1659	.1221	.0739
	8	.0000	.0004	.0046	.0222	.0609	.1144	.1614	.1797	.1623	.1201
	9	.0000	.0001	.0011	.0074	.0271	.0654	.1158	.1597	.1771	.1602
	10	.0000	.0000	.0002	.0020	.0099	.0308	.0686	.1171	.1593	.1762
	11	.0000	.0000	.0000	.0005	.0030	.0120	.0336	.0710	.1185	.1602
	12	.0000	.0000	.0000	.0001	.0008	.0039	.0136	.0355	.0727	.1201
	13	.0000	.0000	.0000	.0000	.0002	.0010	.0045	.0146	.0366	.0739
	14	.0000	.0000	.0000	.0000	.0000	.0002	.0012	.0049	.0150	.0370
	15	.0000	.0000	.0000	.0000	.0000	.0000	.0003	.0013	.0049	.0148
	16	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0003	.0013	.0046
	17	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0011
	18	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0002
	19	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	20	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000

TABLA II

Probabilidades de Poisson†

x	λ 0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0	.9048	.8187	.7408	.6703	.6065	.5488	.4966	.4493	.4066	.3679
1	.0905	.1637	.2222	.2681	.3033	.3293	.3476	.3595	.3659	.3679
2	.0045	.0164	.0333	.0536	.0758	.0988	.1217	.1438	.1647	.1839
3	.0002	.0011	.0033	.0072	.0126	.0198	.0284	.0383	.0494	.0613
4	.0000	.0001	.0002	.0007	.0016	.0030	.0050	.0077	.0111	.0153
5	.0000	.0000	.0000	.0001	.0002	.0004	.0007	.0012	.0020	.0031
6	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0002	.0003	.0005
7	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001

x	λ 1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0
0	.3329	.3012	.2725	.2466	.2231	.2019	.1827	.1653	.1496	.1353
1	.3662	.3614	.3543	.3452	.3347	.3230	.3106	.2975	.2842	.2707
2	.2014	.2169	.2303	.2417	.2510	.2584	.2640	.2678	.2700	.2707
3	.0738	.0867	.0998	.1128	.1255	.1378	.1496	.1607	.1710	.1804
4	.0203	.0260	.0324	.0395	.0471	.0551	.0636	.0723	.0812	.0902
5	.0045	.0062	.0084	.0111	.0141	.0176	.0216	.0260	.0309	.0361
6	.0008	.0012	.0018	.0026	.0035	.0047	.0061	.0078	.0098	.0120
7	.0001	.0002	.0003	.0005	.0008	.0011	.0015	.0020	.0027	.0034
8	.0000	.0000	.0001	.0001	.0001	.0002	.0003	.0005	.0006	.0009
9	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0001	.0001	.0002

x	λ 2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	2.9	3.0
0	.1225	.1108	.1003	.0907	.0821	.0743	.0672	.0608	.0550	.0498
1	.2572	.2438	.2306	.2177	.2052	.1931	.1815	.1703	.1596	.1494
2	.2700	.2681	.2652	.2613	.2565	.2510	.2450	.2384	.2314	.2240
3	.1890	.1966	.2033	.2090	.2138	.2176	.2205	.2225	.2237	.2240
4	.0992	.1082	.1169	.1254	.1336	.1414	.1488	.1557	.1622	.1680
5	.0417	.0476	.0538	.0602	.0668	.0735	.0804	.0872	.0940	.1008
6	.0146	.0174	.0206	.0241	.0278	.0319	.0362	.0407	.0455	.0504
7	.0044	.0055	.0068	.0083	.0099	.0118	.0139	.0163	.0188	.0216
8	.0011	.0015	.0019	.0025	.0031	.0038	.0047	.0057	.0068	.0081
9	.0003	.0004	.0005	.0007	.0009	.0011	.0014	.0018	.0022	.0027
10	.0001	.0001	.0001	.0002	.0002	.0003	.0004	.0005	.0006	.0008
11	.0000	.0000	.0000	.0000	.0000	.0001	.0001	.0001	.0002	.0002
12	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001

† Basado en E. C. Molina, *Poisson's Exponential Binomial Limit*, 1973 Reprint, Robert E. Krieger Publishing Company, Melbourne, Fla., con permiso del editor.

TABLA II (continuación)

	λ									
x	3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9	4.0
0	.0450	.0408	.0369	.0334	.0302	.0273	.0247	.0224	.0202	.0183
1	.1397	.1304	.1217	.1135	.1057	.0984	.0915	.0850	.0789	.0733
2	.2165	.2087	.2008	.1929	.1850	.1771	.1692	.1615	.1539	.1465
3	.2237	.2226	.2209	.2186	.2158	.2125	.2087	.2046	.2001	.1954
4	.1734	.1781	.1823	.1858	.1888	.1912	.1931	.1944	.1951	.1954
5	.1075	.1140	.1203	.1264	.1322	.1377	.1429	.1477	.1522	.1563
6	.0555	.0608	.0662	.0716	.0771	.0826	.0881	.0936	.0989	.1042
7	.0246	.0278	.0312	.0348	.0385	.0425	.0466	.0508	.0551	.0595
8	.0095	.0111	.0129	.0148	.0169	.0191	.0215	.0241	.0269	.0298
9	.0033	.0040	.0047	.0056	.0066	.0076	.0089	.0102	.0116	.0132
10	.0010	.0013	.0016	.0019	.0023	.0028	.0033	.0039	.0045	.0053
11	.0003	.0004	.0005	.0006	.0007	.0009	.0011	.0013	.0016	.0019
12	.0001	.0001	.0001	.0002	.0002	.0003	.0003	.0004	.0005	.0006
13	.0000	.0000	.0000	.0000	.0001	.0001	.0001	.0001	.0002	.0002
14	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001

	λ									
x	4.1	4.2	4.3	4.4	4.5	4.6	4.7	4.8	4.9	5.0
0	.0166	.0150	.0136	.0123	.0111	.0101	.0091	.0082	.0074	.0067
1	.0679	.0630	.0583	.0540	.0500	.0462	.0427	.0395	.0365	.0337
2	.1393	.1323	.1254	.1188	.1125	.1063	.1005	.0948	.0894	.0842
3	.1904	.1852	.1798	.1743	.1687	.1631	.1574	.1517	.1460	.1404
4	.1951	.1944	.1933	.1917	.1898	.1875	.1849	.1820	.1789	.1755
5	.1600	.1633	.1662	.1687	.1708	.1725	.1738	.1747	.1753	.1755
6	.1093	.1143	.1191	.1237	.1281	.1323	.1362	.1398	.1432	.1462
7	.0640	.0686	.0732	.0778	.0824	.0869	.0914	.0959	.1002	.1044
8	.0328	.0360	.0393	.0428	.0463	.0500	.0537	.0575	.0614	.0653
9	.0150	.0168	.0188	.0209	.0232	.0255	.0280	.0307	.0334	.0363
10	.0061	.0071	.0081	.0092	.0104	.0118	.0132	.0147	.0164	.0181
11	.0023	.0027	.0032	.0037	.0043	.0049	.0056	.0064	.0073	.0082
12	.0008	.0009	.0011	.0014	.0016	.0019	.0022	.0026	.0030	.0034
13	.0002	.0003	.0004	.0005	.0006	.0007	.0008	.0009	.0011	.0013
14	.0001	.0001	.0001	.0001	.0002	.0002	.0003	.0003	.0004	.0005
15	.0000	.0000	.0000	.0000	.0001	.0001	.0001	.0001	.0001	.0002

	λ									
x	5.1	5.2	5.3	5.4	5.5	5.6	5.7	5.8	5.9	6.0
0	.0061	.0055	.0050	.0045	.0041	.0037	.0033	.0030	.0027	.0025
1	.0311	.0287	.0265	.0244	.0225	.0207	.0191	.0176	.0162	.0149
2	.0793	.0746	.0701	.0659	.0618	.0580	.0544	.0509	.0477	.0446
3	.1348	.1293	.1239	.1185	.1133	.1082	.1033	.0985	.0938	.0892
4	.1719	.1681	.1641	.1600	.1558	.1515	.1472	.1428	.1383	.1339

TABLA II (continuación)

	λ									
x	5.1	5.2	5.3	5.4	5.5	5.6	5.7	5.8	5.9	6.0
5	.1753	.1748	.1740	.1728	.1714	.1697	.1678	.1656	.1632	.1606
6	.1490	.1515	.1537	.1555	.1571	.1584	.1594	.1601	.1505	.1606
7	.1086	.1125	.1163	.1200	.1234	.1267	.1298	.1326	.1353	.1377
8	.0692	.0731	.0771	.0810	.0849	.0887	.0925	.0962	.0998	.1033
9	.0392	.0423	.0454	.0486	.0519	.0552	.0586	.0620	.0654	.0688
10	.0200	.0220	.0241	.0262	.0285	.0309	.0334	.0359	.0386	.0413
11	.0093	.0104	.0116	.0129	.0143	.0157	.0173	.0190	.0207	.0225
12	.0039	.0045	.0051	.0058	.0065	.0073	.0082	.0092	.0102	.0113
13	.0015	.0018	.0021	.0024	.0028	.0032	.0036	.0041	.0046	.0052
14	.0006	.0007	.0008	.0009	.0011	.0013	.0015	.0017	.0019	.0022
15	.0002	.0002	.0003	.0003	.0004	.0005	.0006	.0007	.0008	.0009
16	.0001	.0001	.0001	.0001	.0001	.0002	.0002	.0002	.0003	.0003
17	.0000	.0000	.0000	.0000	.0000	.0001	.0001	.0001	.0001	.0001

	λ									
x	6.1	6.2	6.3	6.4	6.5	6.6	6.7	6.8	6.9	7.0
0	.0022	.0020	.0018	.0017	.0015	.0014	.0012	.0011	.0010	.0009
1	.0137	.0126	.0116	.0106	.0098	.0090	.0082	.0076	.0070	.0064
2	.0417	.0390	.0364	.0340	.0318	.0296	.0276	.0258	.0240	.0223
3	.0848	.0806	.0765	.0726	.0688	.0652	.0617	.0584	.0552	.0521
4	.1294	.1249	.1205	.1162	.1118	.1076	.1034	.0992	.0952	.0912
5	.1579	.1549	.1519	.1487	.1454	.1420	.1385	.1349	.1314	.1277
6	.1605	.1601	.1595	.1586	.1575	.1562	.1546	.1529	.1511	.1490
7	.1399	.1418	.1435	.1450	.1462	.1472	.1480	.1486	.1489	.1490
8	.1066	.1099	.1130	.1160	.1188	.1215	.1240	.1263	.1284	.1304
9	.0723	.0757	.0791	.0825	.0858	.0891	.0923	.0954	.0985	.1014
10	.0441	.0469	.0498	.0528	.0558	.0588	.0618	.0649	.0679	.0710
11	.0245	.0265	.0285	.0307	.0330	.0353	.0377	.0401	.0426	.0452
12	.0124	.0137	.0150	.0164	.0179	.0194	.0210	.0227	.0245	.0264
13	.0058	.0065	.0073	.0081	.0089	.0098	.0108	.0119	.0130	.0142
14	.0025	.0029	.0033	.0037	.0041	.0046	.0052	.0058	.0064	.0071
15	.0010	.0012	.0014	.0016	.0018	.0020	.0023	.0026	.0029	.0033
16	.0004	.0005	.0005	.0006	.0007	.0008	.0010	.0011	.0013	.0014
17	.0001	.0002	.0002	.0002	.0003	.0003	.0004	.0004	.0005	.0006
18	.0000	.0001	.0001	.0001	.0001	.0001	.0001	.0002	.0002	.0002
19	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0001	.0001

	λ									
x	7.1	7.2	7.3	7.4	7.5	7.6	7.7	7.8	7.9	8.0
0	.0008	.0007	.0007	.0006	.0006	.0005	.0005	.0004	.0004	.0003
1	.0059	.0054	.0049	.0045	.0041	.0038	.0035	.0032	.0029	.0027
2	.0208	.0194	.0180	.0167	.0156	.0145	.0134	.0125	.0116	.0107
3	.0492	.0464	.0438	.0413	.0389	.0366	.0345	.0324	.0305	.0286
4	.0874	.0836	.0799	.0764	.0729	.0696	.0663	.0632	.0602	.0573
5	.1241	.1204	.1167	.1130	.1094	.1057	.1021	.0986	.0951	.0916
6	.1468	.1445	.1420	.1394	.1367	.1339	.1311	.1282	.1252	.1221
7	.1489	.1486	.1481	.1474	.1465	.1454	.1442	.1428	.1413	.1396
8	.1321	.1337	.1351	.1363	.1373	.1382	.1388	.1392	.1395	.1396
9	.1042	.1070	.1096	.1121	.1144	.1167	.1187	.1207	.1224	.1241

TABLA II (continuación)

	λ									
x	7.1	7.2	7.3	7.4	7.5	7.6	7.7	7.8	7.9	8.0
10	.0740	.0770	.0800	.0829	.0858	.0887	.0914	.0941	.0967	.0993
11	.0478	.0504	.0531	.0558	.0585	.0613	.0640	.0667	.0695	.0722
12	.0283	.0303	.0323	.0344	.0366	.0388	.0411	.0434	.0457	.0481
13	.0154	.0168	.0181	.0196	.0211	.0227	.0243	.0260	.0278	.0296
14	.0078	.0086	.0095	.0104	.0113	.0123	.0134	.0145	.0157	.0169
15	.0037	.0041	.0046	.0051	.0057	.0062	.0069	.0075	.0083	.0090
16	.0016	.0019	.0021	.0024	.0026	.0030	.0033	.0037	.0041	.0045
17	.0007	.0008	.0009	.0010	.0012	.0013	.0015	.0017	.0019	.0021
18	.0003	.0003	.0004	.0004	.0005	.0006	.0006	.0007	.0008	.0009
19	.0001	.0001	.0001	.0002	.0002	.0002	.0003	.0003	.0003	.0004
20	.0000	.0000	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0002
21	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0001

	λ									
x	8.1	8.2	8.3	8.4	8.5	8.6	8.7	8.8	8.9	9.0
0	.0003	.0003	.0002	.0002	.0002	.0002	.0002	.0002	.0001	.0001
1	.0025	.0023	.0021	.0019	.0017	.0016	.0014	.0013	.0012	.0011
2	.0100	.0092	.0086	.0079	.0074	.0068	.0063	.0058	.0054	.0050
3	.0269	.0252	.0237	.0222	.0208	.0195	.0183	.0171	.0160	.0150
4	.0544	.0517	.0491	.0466	.0443	.0420	.0398	.0377	.0357	.0337
5	.0882	.0849	.0816	.0784	.0752	.0722	.0692	.0663	.0635	.0607
6	.1191	.1160	.1128	.1097	.1066	.1034	.1003	.0972	.0941	.0911
7	.1378	.1358	.1338	.1317	.1294	.1271	.1247	.1222	.1197	.1171
8	.1395	.1392	.1388	.1382	.1375	.1366	.1356	.1344	.1332	.1318
9	.1256	.1269	.1280	.1290	.1299	.1306	.1311	.1315	.1317	.1318
10	.1017	.1040	.1063	.1084	.1104	.1123	.1140	.1157	.1172	.1186
11	.0749	.0776	.0802	.0828	.0853	.0878	.0902	.0925	.0948	.0970
12	.0505	.0530	.0555	.0579	.0604	.0629	.0654	.0679	.0703	.0728
13	.0315	.0334	.0354	.0374	.0395	.0416	.0438	.0459	.0481	.0504
14	.0182	.0196	.0210	.0225	.0240	.0256	.0272	.0289	.0306	.0324
15	.0098	.0107	.0116	.0126	.0136	.0147	.0158	.0169	.0182	.0194
16	.0050	.0055	.0060	.0066	.0072	.0079	.0086	.0093	.0101	.0109
17	.0024	.0026	.0029	.0033	.0036	.0040	.0044	.0048	.0053	.0058
18	.0011	.0012	.0014	.0015	.0017	.0019	.0021	.0024	.0026	.0029
19	.0005	.0005	.0006	.0007	.0008	.0009	.0010	.0011	.0012	.0014
20	.0002	.0002	.0002	.0003	.0003	.0004	.0004	.0005	.0005	.0006
21	.0001	.0001	.0001	.0001	.0001	.0002	.0002	.0002	.0002	.0003
22	.0000	.0000	.0000	.0000	.0001	.0001	.0001	.0001	.0001	.0001

	λ									
x	9.1	9.2	9.3	9.4	9.5	9.6	9.7	9.8	9.9	10
0	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0000
1	.0010	.0009	.0009	.0008	.0007	.0007	.0006	.0005	.0005	.0005
2	.0046	.0043	.0040	.0037	.0034	.0031	.0029	.0027	.0025	.0023
3	.0140	.0131	.0123	.0115	.0107	.0100	.0093	.0087	.0081	.0076
4	.0319	.0302	.0285	.0269	.0254	.0240	.0226	.0213	.0201	.0189

TABLA II (continuación)

	λ									
x	9.1	9.2	9.3	9.4	9.5	9.6	9.7	9.8	9.9	10
5	.0581	.0555	.0530	.0506	.0483	.0460	.0439	.0418	.0398	.0378
6	.0881	.0851	.0822	.0793	.0764	.0736	.0709	.0682	.0656	.0631
7	.1145	.1118	.1091	.1064	.1037	.1010	.0982	.0955	.0928	.0901
8	.1302	.1286	.1269	.1251	.1232	.1212	.1191	.1170	.1148	.1126
9	.1317	.1315	.1311	.1306	.1300	.1293	.1284	.1274	.1263	.1251
10	.1198	.1210	.1219	.1228	.1235	.1241	.1245	.1249	.1250	.1251
11	.0991	.1012	.1031	.1049	.1067	.1083	.1098	.1112	.1125	.1137
12	.0752	.0776	.0799	.0822	.0844	.0866	.0888	.0908	.0928	.0948
13	.0526	.0549	.0572	.0594	.0617	.0640	.0662	.0685	.0707	.0729
14	.0342	.0361	.0380	.0399	.0419	.0439	.0459	.0479	.0500	.0521
15	.0208	.0221	.0235	.0250	.0265	.0281	.0297	.0313	.0330	.0347
16	.0118	.0127	.0137	.0147	.0157	.0168	.0180	.0192	.0204	.0217
17	.0063	.0069	.0075	.0081	.0088	.0095	.0103	.0111	.0119	.0128
18	.0032	.0035	.0039	.0042	.0046	.0051	.0055	.0060	.0065	.0071
19	.0015	.0017	.0019	.0021	.0023	.0026	.0028	.0031	.0034	.0037
20	.0007	.0008	.0009	.0010	.0011	.0012	.0014	.0015	.0017	.0019
21	.0003	.0003	.0004	.0004	.0005	.0006	.0006	.0007	.0008	.0009
22	.0001	.0001	.0002	.0002	.0002	.0002	.0003	.0003	.0004	.0004
23	.0000	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0002	.0002
24	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0001	.0001

	λ									
x	11	12	13	14	15	16	17	18	19	20
0	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
1	.0002	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
2	.0010	.0004	.0002	.0001	.0000	.0000	.0000	.0000	.0000	.0000
3	.0037	.0018	.0008	.0004	.0002	.0001	.0000	.0000	.0000	.0000
4	.0102	.0053	.0027	.0013	.0006	.0003	.0001	.0001	.0000	.0000
5	.0224	.0127	.0070	.0037	.0019	.0010	.0005	.0002	.0001	.0001
6	.0411	.0255	.0152	.0087	.0048	.0026	.0014	.0007	.0004	.0002
7	.0646	.0437	.0281	.0174	.0104	.0060	.0034	.0018	.0010	.0005
8	.0888	.0655	.0457	.0304	.0194	.0120	.0072	.0042	.0024	.0013
9	.1085	.0874	.0661	.0473	.0324	.0213	.0135	.0083	.0050	.0029
10	.1194	.1048	.0859	.0663	.0486	.0341	.0230	.0150	.0095	.0058
11	.1194	.1144	.1015	.0844	.0663	.0496	.0355	.0245	.0164	.0106
12	.1094	.1144	.1099	.0984	.0829	.0661	.0504	.0368	.0259	.0176
13	.0926	.1056	.1099	.1060	.0956	.0814	.0658	.0509	.0378	.0271
14	.0728	.0905	.1021	.1060	.1024	.0930	.0800	.0655	.0514	.0387
15	.0534	.0724	.0885	.0989	.1024	.0992	.0906	.0786	.0650	.0516
16	.0367	.0543	.0719	.0866	.0960	.0992	.0963	.0884	.0772	.0646
17	.0237	.0383	.0550	.0713	.0847	.0934	.0963	.0936	.0863	.0760
18	.0145	.0256	.0397	.0554	.0706	.0830	.0909	.0936	.0911	.0844
19	.0084	.0161	.0272	.0409	.0557	.0699	.0814	.0887	.0911	.0888
20	.0046	.0097	.0177	.0286	.0418	.0559	.0692	.0798	.0866	.0888
21	.0024	.0055	.0109	.0191	.0299	.0426	.0560	.0684	.0783	.0846
22	.0012	.0030	.0065	.0121	.0204	.0310	.0433	.0560	.0676	.0769
23	.0006	.0016	.0037	.0074	.0133	.0216	.0320	.0438	.0559	.0669
24	.0003	.0008	.0020	.0043	.0083	.0144	.0226	.0328	.0442	.0557

TABLA II (continuación)

	λ									
x	11	12	13	14	15	16	17	18	19	20
25	.0001	.0004	.0010	.0024	.0050	.0092	.0154	.0237	.0336	.0446
26	.0000	.0002	.0005	.0013	.0029	.0057	.0101	.0164	.0246	.0343
27	.0000	.0001	.0002	.0007	.0016	.0034	.0063	.0109	.0173	.0254
28	.0000	.0000	.0001	.0003	.0009	.0019	.0038	.0070	.0117	.0181
29	.0000	.0000	.0001	.0002	.0004	.0011	.0023	.0044	.0077	.0125
30	.0000	.0000	.0000	.0001	.0002	.0006	.0013	.0026	.0049	.0083
31	.0000	.0000	.0000	.0000	.0001	.0003	.0007	.0015	.0030	.0054
32	.0000	.0000	.0000	.0000	.0001	.0001	.0004	.0009	.0018	.0034
33	.0000	.0000	.0000	.0000	.0000	.0001	.0002	.0005	.0010	.0020
34	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0002	.0006	.0012
35	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0007
36	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0002	.0004
37	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0002
38	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001
39	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001

TABLA III

Distribución normal estándar

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4988
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990

También, para z = 4.0, 5.0, y 6.0, las probabilidades son 0.49997, 0.4999997 y 0.499999999.

TABLA VI

Valores de $f_{0.05,\nu_1,\nu_2}$†

ν_1 = Grados de libertad del numerador

ν_2 = Grados de libertad del denominador	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161	200	216	225	230	234	237	239	241	242	244	246	248	249	250	251	252	253	254
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.5	19.5	19.5	19.5	19.5	19.5
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.37
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

† Reproducido de M. Merrington and C. M. Thompson, "Tables of percentage points of the inverted beta (*F*) distribution," *Biometrika*, vol. 33 (1943), con permiso de los fideicomisarios de *Biometrika*.

TABLA VI (continuación)

Valores de $f_{0.01,\nu_1,\nu_2}$

ν_1 = Grados de libertad del numerador

ν_2 = Grados de libertad del denominador	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	4,052	5,000	5,403	5,625	5,764	5,859	5,928	5,982	6,023	6,056	6,106	6,157	6,209	6,235	6,261	6,287	6,313	6,339	6,366
2	98.5	99.0	99.2	99.2	99.3	99.3	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.5	99.5	99.5	99.5	99.5	99.5
3	34.1	30.8	29.5	28.7	28.2	27.9	27.7	27.5	27.3	27.2	27.1	26.9	26.7	26.6	26.5	26.4	26.3	26.2	26.1
4	21.2	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.7	14.5	14.4	14.2	14.0	13.9	13.8	13.7	13.7	13.6	13.5
5	16.3	13.3	12.1	11.4	11.0	10.7	10.5	10.3	10.2	10.1	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	13.7	10.9	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
7	12.2	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.07	5.99	6.91	5.82	5.74	5.65
8	11.3	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9	10.6	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	10.0	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17
14	8.86	6.51	5.56	5.04	4.70	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75
17	8.40	6.11	5.19	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	8.19	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	7.77	5.57	4.68	4.18	3.86	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.62	2.53	2.45	2.36	2.27	2.17
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00

TABLA VII

Factoriales

n	$n!$	$\log n!$
0	1	0.0000
1	1	0.0000
2	2	0.3010
3	6	0.7782
4	24	1.3802
5	120	2.0792
6	720	2.8573
7	5,040	3.7024
8	40,320	4.6055
9	362,880	5.5598
10	3,628,800	6.5598
11	39,916,800	7.6012
12	479,001,600	8.6803
13	6,227,020,800	9.7943
14	87,178,291,200	10.9404
15	1,307,674,368,000	12.1165

Coeficientes binomiales

n	$\binom{n}{0}$	$\binom{n}{1}$	$\binom{n}{2}$	$\binom{n}{3}$	$\binom{n}{4}$	$\binom{n}{5}$	$\binom{n}{6}$	$\binom{n}{7}$	$\binom{n}{8}$	$\binom{n}{9}$	$\binom{n}{10}$
0	1										
1	1	1									
2	1	2	1								
3	1	3	3	1							
4	1	4	6	4	1						
5	1	5	10	10	5	1					
6	1	6	15	20	15	6	1				
7	1	7	21	35	35	21	7	1			
8	1	8	28	56	70	56	28	8	1		
9	1	9	36	84	126	126	84	36	9	1	
10	1	10	45	120	210	252	210	120	45	10	1
11	1	11	55	165	330	462	462	330	165	55	11
12	1	12	66	220	495	792	924	792	495	220	66
13	1	13	78	286	715	1287	1716	1716	1287	715	286
14	1	14	91	364	1001	2002	3003	3432	3003	2002	1001
15	1	15	105	455	1365	3003	5005	6435	6435	5005	3003
16	1	16	120	560	1820	4368	8008	11440	12870	11440	8008
17	1	17	136	680	2380	6188	12376	19448	24310	24310	19448
18	1	18	153	816	3060	8568	18564	31824	43758	48620	43758
19	1	19	171	969	3876	11628	27132	50388	75582	92378	92378
20	1	20	190	1140	4845	15504	38760	77520	125970	167960	184756

TABLA IX

Valores de r_p para $\alpha = 0.01$†

g.l. \ p	2	3	4	5	6	7	8	9	10
1	90.02								
2	14.04	14.04							
3	8.26	8.32	8.32						
4	6.51	6.68	6.74	6.76					
5	5.70	5.90	5.99	6.04	6.07				
6	5.24	5.44	5.55	5.62	5.66	5.68			
7	4.95	5.15	5.26	5.33	5.38	5.42	5.44		
8	4.74	4.94	5.06	5.13	5.19	5.23	5.26	5.28	
9	4.60	4.79	4.91	4.99	5.04	5.09	5.12	5.14	5.16
10	4.48	4.67	4.79	4.88	4.93	4.98	5.01	5.04	5.06
11	4.39	4.58	4.70	4.78	4.84	4.89	4.92	4.95	4.97
12	4.32	4.50	4.62	4.71	4.77	4.81	4.85	4.88	4.91
13	4.26	4.44	4.56	4.64	4.71	4.75	4.79	4.82	4.85
14	4.21	4.39	4.51	4.59	4.66	4.70	4.74	4.77	4.80
15	4.17	4.34	4.46	4.55	4.61	4.66	4.70	4.73	4.76
16	4.13	4.31	4.43	4.51	4.57	4.62	4.66	4.70	4.72
17	4.10	4.27	4.39	4.47	4.54	4.59	4.63	4.66	4.69
18	4.07	4.25	4.36	4.45	4.51	4.56	4.60	4.64	4.66
19	4.05	4.22	4.33	4.42	4.48	4.53	4.57	4.61	4.64
20	4.02	4.20	4.31	4.40	4.46	4.51	4.55	4.59	4.62
24	3.96	4.13	4.24	4.32	4.39	4.44	4.48	4.52	4.55
30	3.89	4.06	4.17	4.25	4.31	4.36	4.41	4.45	4.48
40	3.82	3.99	4.10	4.18	4.24	4.29	4.33	4.38	4.41
60	3.76	3.92	4.03	4.11	4.18	4.23	4.37	4.31	4.34
120	3.70	3.86	3.97	4.04	4.11	4.16	4.20	4.24	4.27
∞	3.64	3.80	3.90	3.98	4.04	4.09	4.13	4.17	4.21

† Esta tabla se reproduce de H. L. Harter, "Critical Values for Duncan's New Multiple Range Test". Contiene algunos valores corregidos para reemplazar los que brinda D. B. Duncan en "Multiple Range and Multiple F Tests", *Biometrics*, vol. 11 (1955). La tabla de arriba se reproduce bajo permiso del autor de la Biometric Society.

TABLA IX (continuación)

Valores de r_p para $\alpha = 0.05$

g.l. \ p	2	3	4	5	6	7	8	9	10
1	17.97								
2	6.09	6.09							
3	4.50	4.52	4.52						
4	3.93	4.01	4.03	4.03					
5	3.64	3.75	3.80	3.81	3.81				
6	3.46	3.59	3.65	3.68	3.69	3.70			
7	3.34	3.48	3.55	3.59	3.61	3.62	3.63		
8	3.26	3.40	3.48	3.52	3.55	3.57	3.57	3.58	
9	3.20	3.34	3.42	3.47	3.50	3.52	3.54	3.54	3.55
10	3.15	3.29	3.38	3.43	3.47	3.49	3.51	3.52	3.52
11	3.11	3.26	3.34	3.40	3.44	3.46	3.48	3.49	3.50
12	3.08	3.23	3.31	3.37	3.41	3.44	3.46	3.47	3.48
13	3.06	3.20	3.29	3.35	3.39	3.42	3.46	3.46	3.47
14	3.03	3.18	3.27	3.33	3.37	3.40	3.43	3.44	3.46
15	3.01	3.16	3.25	3.31	3.36	3.39	3.41	3.43	3.45
16	3.00	3.14	3.23	3.30	3.34	3.38	3.40	3.42	3.44
17	2.98	3.13	3.22	3.28	3.33	3.37	3.39	3.41	3.43
18	2.97	3.12	3.21	3.27	3.32	3.36	3.38	3.40	3.42
19	2.96	3.11	3.20	3.26	3.31	3.35	3.38	3.40	3.41
20	2.95	3.10	3.19	3.25	3.30	3.34	3.37	3.39	3.41
24	2.92	3.07	3.16	3.23	3.28	3.31	3.35	3.37	3.39
30	2.89	3.03	3.13	3.20	3.25	3.29	3.32	3.35	3.37
40	2.86	3.01	3.10	3.17	3.22	3.27	3.30	3.33	3.35
60	2.83	2.98	3.07	3.14	3.20	3.24	3.28	3.31	3.33
120	2.80	2.95	3.04	3.12	3.17	3.22	3.25	3.29	3.31
∞	2.77	2.92	3.02	3.09	3.15	3.19	3.23	3.27	3.29

TABLA X

Valores críticos para la prueba de rangos con signo†

n	$T_{0.10}$	$T_{0.05}$	$T_{0.02}$	$T_{0.01}$
4				
5	1			
6	2	1		
7	4	2	0	
8	6	4	2	0
9	8	6	3	2
10	11	8	5	3
11	14	11	7	5
12	17	14	10	7
13	21	17	13	10
14	26	21	16	13
15	30	25	20	16
16	36	30	24	19
17	41	35	28	23
18	47	40	33	28
19	54	46	38	32
20	60	52	43	37
21	68	59	49	43
22	75	66	56	49
23	83	73	62	55
24	92	81	69	61
25	101	90	77	68

† De F. Wilcoxon and R. A. Wilcox, *Some Rapid Approximate Statistical Procedures*, American Cyanamid Company, Pearl River, N.Y., 1964. Reproducido con permiso de American Cyanamid Company.

TABLA XI

Valores críticos para la prueba U†

Valores de $U_{0.10}$

n_1 \ n_2	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2				0	0	0	1	1	1	1	2	2	3	3
3		0	0	1	2	2	3	4	4	5	5	6	7	7
4		0	1	2	3	4	5	6	7	8	9	10	11	12
5	0	1	2	4	5	6	8	9	11	12	13	15	16	18
6	0	2	3	5	7	8	10	12	14	16	17	19	21	23
7	0	2	4	6	8	11	13	15	17	19	21	24	26	28
8	1	3	5	8	10	13	15	18	20	23	26	28	31	33
9	1	4	6	9	12	15	18	21	24	27	30	33	36	39
10	1	4	7	11	14	17	20	24	27	31	34	37	41	44
11	1	5	8	12	16	19	23	27	31	34	38	42	46	50
12	2	5	9	13	17	21	26	30	34	38	42	47	51	55
13	2	6	10	15	19	24	28	33	37	42	47	51	56	61
14	3	7	11	16	21	26	31	36	41	46	51	56	61	66
15	3	7	12	18	23	28	33	39	44	50	55	61	66	72

Valores de $U_{0.05}$

n_1 \ n_2	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2							0	0	0	0	1	1	1	1
3				0	1	1	2	2	3	3	4	4	5	5
4			0	1	2	3	4	4	5	6	7	8	9	10
5		0	1	2	3	5	6	7	8	9	11	12	13	14
6		1	2	3	5	6	8	10	11	13	14	16	17	19
7		1	3	5	6	8	10	12	14	16	18	20	22	24
8	0	2	4	6	8	10	13	15	17	19	22	24	26	29
9	0	2	4	7	10	12	15	17	20	23	26	28	31	34
10	0	3	5	8	11	14	17	20	23	26	29	30	36	39
11	0	3	6	9	13	16	19	23	26	30	33	37	40	44
12	1	4	7	11	14	18	22	26	29	33	37	41	45	49
13	1	4	8	12	16	20	24	28	30	37	41	45	50	54
14	1	5	9	13	17	22	26	31	36	40	45	50	55	59
15	1	5	10	14	19	24	29	34	39	44	49	54	59	64

† Esta tabla se basa en D. Auble, "Extended Tables for the Mann-Whitney Statistics", en *Bulletin of the Institute of Educational Research at Indiana University*, vol. 1, 1953. Con permiso del autor.

TABLA XI (continuación)

Valores de $U_{0.02}$

n_1 \ n_2	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2												0	0	0
3						0	0	1	1	1	2	2	2	3
4				0	1	1	2	3	3	4	5	5	6	7
5			0	1	2	3	4	5	6	7	8	9	10	11
6			1	2	3	4	6	7	8	9	11	12	13	15
7		0	1	3	4	6	7	9	11	12	14	16	17	19
8		0	2	4	6	7	9	11	13	15	17	20	22	24
9		1	3	5	7	9	11	14	16	18	21	23	26	28
10		1	3	6	8	11	13	16	19	22	24	27	30	33
11		1	4	7	9	12	15	18	22	25	28	31	34	37
12		2	5	8	11	14	17	21	24	28	31	35	38	42
13	0	2	5	9	12	16	20	23	27	31	35	39	43	47
14	0	2	6	10	13	17	22	26	30	34	38	43	47	51
15	0	3	7	11	15	19	24	28	33	37	42	47	51	56

Valores de $U_{0.01}$

n_1 \ n_2	3	4	5	6	7	8	9	10	11	12	13	14	15
3							0	0	0	1	1	1	2
4				0	0	1	1	2	2	3	3	4	5
5			0	1	1	2	3	4	5	6	7	7	8
6		0	1	2	3	4	5	6	7	9	10	11	12
7		0	1	3	4	6	7	9	10	12	13	15	16
8		1	2	4	6	7	9	11	13	15	17	18	20
9	0	1	3	5	7	9	11	13	16	18	20	22	24
10	0	2	4	6	9	11	13	16	18	21	24	26	29
11	0	2	5	7	10	13	16	18	21	24	27	30	33
12	1	3	6	9	12	15	18	21	24	27	31	34	37
13	1	3	7	10	13	17	20	24	27	31	34	38	42
14	1	4	7	11	15	18	22	26	30	34	38	42	46
15	2	5	8	12	16	20	24	29	33	37	42	46	51

CAPÍTULO 2

2.5 (a) $\{6, 8, 9\}$; (b) $\{8\}$; (c) $\{1, 2, 3, 4, 5, 8\}$; (d) $\{1, 5\}$; (e) $\{2, 4, 8\}$; (f) $\emptyset$.

2.7 (a) {Auto 5, Auto 6, Auto 7, Auto 8}; (b) {Auto 2, Auto 4, Auto 5, Auto 7}; (c) {Auto 1, Auto 8}; (d) {Auto 3, Auto 4, Auto 7, Auto 8}.

2.9 (a) La casa tiene menos de tres baños. (b) La casa no tiene una chimenea. (c) La casa no cuesta más de \$100.000. (d) La casa no es nueva. (e) La casa tiene tres o más baños y una chimenea. (f) La casa tiene tres o más baños y cuesta más de \$100,000. (g) La casa cuesta más de \$100,000 pero no tiene chimenea. (h) La casa es nueva y cuesta más de \$100,000. (i) La casa es nueva o cuesta más de \$100,000. (j) La casa tiene tres o más baños y/o una chimenea. (k) La casa tiene tres o más baños y/o cuesta más de \$100,000. (l) La casa es nueva y cuesta más de \$100,000.

2.11 (a) (H, 1), (H, 2), (H, 3), (H, 4), (H, 5), (H, 6) (T, H, H), (T, H, T), (T, T, H) y (T, T, T); (b) (H, 1), (H, 2), (H, 3), (H, 4), (H, 5), (H, 6), (T, H, T) y (T, T, H); (c) (H, 5), (H, 6), (T, H, T), (T, T, H) y (T, T, T).

2.13 (a) 5^{k-1}; (b) $\dfrac{5^k - 1}{4}$.

2.15 $\{(x, y) | (x - 2)^2 + (y + 3)^2 \leq 9\}$.

2.17 (a) El evento de que una conductora tenga seguro de responsabilidad civil. (b) El evento de que una conductora no tenga seguro contra accidentes. (c) El evento de que una conductora tenga seguro de responsabilidad civil o seguro contra accidentes, pero no ambos. (d) El evento de que una conductora no tenga ambas clases de seguro.

2.19 (a) región 5; (b) regiones 1 y 2 juntas; (c) regiones 3, 5 y 6 juntas; (d) regiones 1, 3, 4 y 6 juntas.

2.21 Las cifras son inconsistentes y se deben poner en duda los resultados del estudio.

2.35 (a) Permisible; (b) no permisible porque la suma de las probabilidades excede a 1; (c) permisible; (d) no permisible porque $P(E)$ es negativa; (e) no permisible porque la suma de las probabilidades es menor que 1.

2.37 (a) La probabilidad de que no pasará no puede ser negativa. (b) $0.77 + 0.08 = 0.85 \neq 0.95$; (c) $0.12 + 0.25 + 0.36 + 0.14 + 0.09 + 0.07 = 1.03 > 1$; (d) $0.08 + 0.21 + 0.29 + 0.40 = 0.98 < 1$.

2.39 (a) 0.29; (b) 0.80; (c) 0.63; (d) 0.71.

2.41 (a) $\frac{3}{8}$; (b) $\frac{1}{4}$; (c) $\frac{1}{10}$; (d) $\frac{1}{10}$; (e) $\frac{11}{40}$.

2.43 $\frac{20}{221}$.

2.45 (a) $\frac{25}{108}$; (b) $\frac{25}{162}$; (c) $\frac{25}{648}$; (d) $\frac{25}{1296}$.

2.47 (a) $P(A \cup B)$ es menor que $P(A)$. (b) $P(A \cap B)$ excede a $P(A)$. (c) $P(A \cup B)$ excede a 1.

2.49 0.34.

2.51 $\frac{13}{36}$.

2.53 0.94.

2.55 (a) 3 a 2; (b) 11 a 5; (c) 7 a 2 en contra

2.57 5 a 3.

2.75 (a) $\frac{3}{5}$; (b) $\frac{7}{10}$; (c) $\frac{1}{5}$; (d) $\frac{3}{10}$; (e) $\frac{1}{3}$; (f) $\frac{3}{7}$.

2.77 $\frac{17}{18}$.

2.79 $\frac{13}{16}$.

2.81 $\frac{1}{3}$.

2.83 0.44.

2.85 (a) 0.096; (b) 0.048; (c) 0.0512; (d) 0.76.

2.87 (a) Los eventos son independientes por parejas; (b) los eventos no son independientes.

2.89 (a) 0.1406; (b) 0.1198.

2.91 0.7176.

2.93 $\frac{1}{91}$.

2.97 0.76.

2.99 0.5684.

2.101 (a) 0.0944; (b) 0.8051.

2.103 (a) 0.032; (b) 0.09375; (c) 0.625.

2.105 (a) 0.6757.

CAPÍTULO 3

3.1 (a) No, porque $f(4)$ es negativa; (b) sí; (c) no, porque la suma de las probabilidades es menor que 1.

3.5 $0 < k < 1$.

3.9 (a) No, porque $F(4)$ excede a 1; (b) no, porque $F(2)$ es menor que $F(1)$; (c) sí.

3.11 $$F(x) = \begin{cases} 0 & \text{para } x < 1 \\ \frac{1}{15} & \text{para } 1 \leqq x < 2 \\ \frac{3}{15} & \text{para } 2 \leqq x < 3 \\ \frac{6}{15} & \text{para } 3 \leqq x < 4 \\ \frac{10}{15} & \text{para } 4 \leqq x < 5 \\ 1 & \text{para } x \geqq 5 \end{cases}$$

3.13 (a) $\frac{3}{4}$; (b) $\frac{1}{4}$; (c) $\frac{1}{2}$; (d) $\frac{3}{4}$; (e) $\frac{1}{2}$; (f) $\frac{1}{4}$.

3.17 (a) $f(3) = \frac{1}{6}$, $f(4) = \frac{1}{6}$, $f(5) = \frac{1}{3}$, $f(6) = \frac{1}{6}$ y $f(7) = \frac{1}{6}$;

(b) $$F(z) = \begin{cases} 0 & \text{para } z < 3 \\ \frac{1}{6} & \text{para } 3 \leqq z < 4 \\ \frac{1}{3} & \text{para } 4 \leqq z < 5 \\ \frac{2}{3} & \text{para } 5 \leqq z < 6 \\ \frac{5}{6} & \text{para } 6 \leqq z < 7 \\ 1 & \text{para } z \geqq 7 \end{cases}$$

3.19 $$F(x) = \begin{cases} 0 & \text{para } x < 0 \\ \frac{1}{27} & \text{para } 0 \leqq x < 1 \\ \frac{7}{27} & \text{para } 1 \leqq x < 2 \\ \frac{19}{27} & \text{para } 2 \leqq x < 3 \\ 1 & \text{para } x \geqq 3 \end{cases}$$

(a) $\frac{20}{27}$; (b) $\frac{8}{27}$.

4.49 $\frac{15}{44}$.

4.51 $\mu = 0$ y $\sigma^2 = 7.2$.

4.53 (a) 0.631; (b) 0.553.

4.55 (a) Entre 0.230 y 0.290; (b) entre 0.200 y 0.320.

4.57 8.

4.59 0.

4.61 Por ejemplo, $f(0, 0) = \frac{1}{6}$ y $g(0)h(0) = \frac{1}{24}$, de manera que $f(0, 0) \neq g(0)h(0)$.

4.63 (c) $E(e^{t_1X+t_2Y}) = \dfrac{1}{(1 - t_1)(1 - t_2)}$, $E(XY) = 1$, $E(X) = 1$, $E(Y) = 1$ y $\text{cov}(X, Y) = 0$.

4.65 (a) 143; (b) 54.

4.69 75.

4.71 $\mu_{X|-1} = \frac{3}{5}$ y $\sigma^2_{X|-1} = \frac{16}{25}$.

4.73 $\mu_{Y|1/4} = \frac{11}{9}$ y $\sigma^2_{Y|1/4} = \frac{23}{81}$.

4.77 0.0224.

4.79 (a) $\mu = 0.74$ y $\sigma = 0.68$; (b) $\mu = 1.91$ y $\sigma = 1.05$.

4.81 0.8.

4.83 2.95 minutos.

CAPÍTULO 5

5.11 $\mu'_2 = \mu'_{(2)} + \mu'_{(1)}$, $\mu'_3 = \mu'_{(3)} + 3\mu'_{(2)} + \mu'_{(1)}$ y $\mu'_4 = \mu'_{(4)} + 6\mu'_{(3)} + 7\mu'_{(2)} + \mu'_{(1)}$.

5.13 (a) $F_X(t) = 1 - \theta + \theta t$; (b) $F_X(t) = [1 + \theta(t - 1)]^n$.

5.15 (a) $\alpha_3 = 0$ cuando $\theta = \frac{1}{2}$; (b) $\alpha_3 \to 0$ cuando $n \to \infty$.

5.17 0.0086.

5.19 (a) 0.3025; (b) 0.3025.

5.21 (a) 0.2205; (b) 0.2206.

5.23 0.2041.

5.25 0.9222.

5.27 0.0754.

5.29 (a) 0.0538.

5.33 $f(y) = \binom{y + k - 1}{k - 1}\theta^k(1 - \theta)^y$ para $y = 0, 1, 2, \ldots$.

5.43 $h(0; 4, 9, 5) = \frac{1}{126}$, $h(1; 4, 9, 5) = \frac{20}{126}$, $h(2; 4, 9, 5) = \frac{60}{126}$, $h(3; 4, 9, 5) = \frac{40}{126}$ y $h(4; 4, 9, 5) = \frac{5}{126}$.

5.47 (a) 0.0060; (b) 0.0076.

5.53 $\mu_2 = \lambda$, $\mu_3 = \lambda$ y $\mu_4 = \lambda + 3\lambda^2$.

5.55 (a) 0.1298; (b) 0.1101.

5.57 (a) 0.0180; (b) 0.0180.

6.41 (a) $\mu = 0.2$; (b) 0.3164.

6.45 $\mu_3 = 0$ y $\mu_4 = 3\sigma^4$.

6.55 (a) 0.1271; (b) 0.6406; (c) 0.1413; (d) 0.5876.

6.57 (a) $z = 1.92$; (b) $z = 2.22$; (c) $z = 1.12$; (d) $z = \pm 1.44$.

6.59 (a) 0.6826; (b) 0.9544; (c) 0.9984; (d) 0.99994.

6.61 (a) 0.6280; (b) 0.6279.

6.63 (a) 0.0668; (b) 0.2860; (c) 0.6490.

6.65 6.094 onzas.

6.67 (a) No; (b) sí; (c) no.

6.71 0.1446.

6.73 (a) 0.24; (b) 0.49; (c) 0.96.

6.77 $\sigma_1 = 6$, $\sigma_2 = 3$ y $\rho = -\frac{1}{2}$.

6.79 $\rho_{UV} = \dfrac{\sigma_1^2 - \sigma_2^2}{\sqrt{(\sigma_1^2 + \sigma_2^2)^2 - 4\rho^2\sigma_1^2\sigma_2^2}}$.

6.81 (a) 0.2990; (b) 0.18.

6.83 (a) 14.5 libras; (b) 23.625 pulgadas.

CAPÍTULO 7

7.1 (a) $G(y) = 1 - e^{-y}$ para $y > 0$ y $G(y) = 0$ en cualquier otra parte; (b) $g(y) = e^{-y}$ para $y > 0$ y $g(y) = 0$ en cualquier otra parte.

7.3 $g(y) = 2y$ para $0 < y < 1$ y $g(y) = 0$ en cualquier otra parte.

7.5 (a) $f(y) = \dfrac{1}{\theta_1 - \theta_2}\cdot(e^{-y/\theta_1} - e^{-y/\theta_2})$ para $y > 0$ y $f(y) = 0$ en cualquier otra parte;

(b) $f(y) = \dfrac{1}{\theta^2}\cdot ye^{-y/\theta}$ para $y > 0$ y $f(y) = 0$ en cualquier otra parte.

7.7 (a) $F(y) = 0$; (b) $F(y) = \frac{1}{2}y^2$; (c) $F(y) = 1 - \frac{1}{2}(2-y)^2$; (d) $F(y) = 1$. También, $f(y) = 0$ para $y \leqq 0$, $f(y) = y$ para $0 < y \leqq 1$, $f(y) = 2 - y$ para $1 < y < 2$ y $f(y) = 0$ para $y \geqq 2$.

7.9 $g(v) = e^{-v}$ para $v > 0$ y $g(v) = 0$ en cualquier otra parte.

7.11 $g(z) = \frac{1}{25}\cdot(20\cdot\ln 2 - 10)$ para $0 < z \leqq 5$, $g(z) = \frac{1}{25}\cdot\left(2z - 20 - 20\cdot\ln\frac{z}{10}\right)$ para $5 < z < 10$, y $g(z) = 0$ en cualquier otra parte.

7.13 $g(y) = \frac{9}{11}\cdot y^2$ para $0 < y \leqq 1$, $g(y) = \dfrac{3(2-y)(7y-4)}{22}$ para $1 < y < 2$ y $g(y) = 0$ en cualquier otra parte.

7.15 $h(0) = \frac{1}{3}$ y $h(1) = \frac{2}{3}$.

7.17 $g(y) = \theta(1-\theta)^{-(1+y)/5}$ para $y = -1, -6, -11, \ldots$.

7.21 $g(y) = \dfrac{1}{\sigma\sqrt{2\pi}}\cdot\dfrac{1}{y}\cdot e^{-\frac{1}{2}\left(\frac{\ln y - \mu}{\sigma}\right)^2}$ para $y > 0$ y $g(y) = 0$ en cualquier otra parte.

ESTADÍSTICA MATEMÁTICA CON APLICACIONES

Sexta edición

John E. Freund, Irwin Miller y Marylees Miller

Es un texto clásico que proporciona sólidos fundamentos matemáticos de la estadística. Para utilizarlo se requieren conocimientos de Cálculo diferencial e integral.

En el libro se desarrolla claramente la teoría mediante la demostración y ejemplificación de los teoremas importantes; la exposición de los temas es clara y de buen nivel matemático. Esta obra contiene un mayor número de ejercicios, aplicaciones y problemas para resolver con la computadora, y enseña al alumno a modelar situaciones donde interviene la incertidumbre. Los autores dan un tratamiento cuidadoso a la probabilidad.

Entre los temas nuevos se encuentran: el modelo de análisis de varianza en dos sentidos con interacción, una revisión de las comparaciones múltiples y, en los apéndices, las propiedades de las funciones de distribución y densidad de probabilidad especiales.

ISBN 970-17-0389-8
90000
9 789701 703892

Pearson Educación

Visítenos en:
www.pearson.com.mx