SAH

Introducción a la econometría

Un enfoque moderno

APENDICE B

4a. edición

Jeffrey M. Wooldridge

Michigan State University

Traducción

Ma. del Carmen Enriqueta Hano Roa

Érika M. Jasso Hernan D´Borneville

Traductoras profesionales

Revisión técnica

Roberto Palma Pacheco

Centro de Alta Dirección en Economía y Negocios

Universidad Anáhuac

Domingo Rodríguez Benavides

Facultad de Economía

Universidad Nacional Autónoma de México

319855

CENGAGE Learning

Australia • Brasil • Corea • España • Estados Unidos • Japón • México • Reino Unido • Singapur

Apéndice B

Fundamentos de probabilidad

En este apéndice se hará un repaso de los conceptos clave de la probabilidad. Los apéndices B y C pretenden servir como repaso, no como sustitutos de un curso de probabilidad y estadística. Sin embargo, en ellos se revisan todos los conceptos de probabilidad y estadística que se usan en este libro.

La probabilidad por sí misma es de interés para estudiantes de negocios, economía y otras ciencias sociales. Considere, por ejemplo, el caso de una aerolínea que tiene que decidir cuántas reservaciones aceptar para un vuelo en el que hay 100 asientos disponibles. Si hay menos de 100 personas que quieran hacer una reservación, todas serán aceptadas. Pero, ¿qué pasa si hay más de 100 personas que quieran una reservación? Una solución segura es aceptar como máximo 100 reservaciones. Sin embargo, dado que hay personas que hacen una reservación y después no la utilizan, existe la posibilidad de que aun cuando se hayan hecho 100 reservaciones, el avión no se llene. Esto ocasiona pérdidas a la aerolínea. Otra solución puede ser aceptar más de 100 reservaciones esperando que no todas sean utilizadas, de manera que la cantidad final de pasajeros sea lo más cercana posible a 100. Con esta política, la aerolínea corre el riesgo de tener que compensar a las personas que, debido al exceso de reservaciones, no puedan ser aceptadas para el vuelo.

La pregunta natural que surge en este contexto es: ¿se puede determinar la cantidad óptima (o mejor) de reservaciones que deba hacer la aerolínea? Este no es un problema trivial. Sin embargo, dada cierta información (sobre los costos de la aerolínea y qué tan frecuente es que las personas utilicen su reservación), puede emplearse la probabilidad básica para encontrar la solución a este caso.

B.1 Variables aleatorias y sus distribuciones de probabilidad

Suponga que lanza una moneda 10 veces y se cuenta la cantidad de ellas en que se obtiene cara. Esto es un ejemplo de un **experimento**. En general, un experimento es un procedimiento que puede, al menos en teoría, repetirse una cantidad infinita de veces y que tiene un conjunto bien definido de resultados. Lanzar una moneda puede, en principio, repetirse una y otra vez. Antes de lanzarla 10 veces, ya se sabe que la cantidad de caras que puede obtenerse será un número desde 0 hasta 10, de manera que los resultados de este experimento están bien definidos.

Una **variable aleatoria** es aquella que toma un valor numérico que será determinado por un experimento. En el ejemplo del lanzamiento de la moneda, la cantidad de caras en 10 lanzamientos es un ejemplo de una variable aleatoria. Antes de lanzar la moneda 10 veces, no se sabe en cuántas de ellas se obtendrá cara. Una vez lanzada 10 veces y contado el número de caras, se obtiene el valor de la variable aleatoria en este particular ensayo del experimento. En otro ensayo puede obtenerse otro resultado.

En el ejemplo antes citado de las reservaciones de la aerolínea, la cantidad de personas que utiliza su reservación es una variable aleatoria: antes de un determinado vuelo no se sabe cuántas personas utilizan su reservación.

Para analizar los datos recolectados, en negocios y en ciencias sociales, es importante tener un conocimiento básico sobre las variables aleatorias y sus propiedades. De acuerdo con la convención usual en probabilidad y estadística, en los apéndices B y C las variables aleatorias se denotarán empleando letras mayúsculas, por lo general W, X, Y y Z, y los valores particulares de las variables aleatorias empleando las correspondientes letras minúsculas, w, x, y y z. Por ejemplo, en el experimento del lanzamiento de una moneda, sea X la cantidad de caras obtenidas en los 10 lanzamientos de la misma. Entonces, X no está relacionada con ningún valor determinado, pero se sabe que X toma uno de los valores del conjunto $\{0, 1, 2, \ldots, 10\}$. Un valor particular es, por ejemplo, $x = 6$.

Para colecciones grandes de variables aleatorias se usan subíndices. Por ejemplo, si se obtienen los ingresos del último año de 20 hogares de Estados Unidos, elegidos aleatoriamente, estas variables aleatorias se denotarán $X_1, X_2, \ldots, X_{20}$; y los valores particulares se denotarán $x_1, x_2, \ldots, x_{20}$.

Como se establece en la definición, las variables aleatorias siempre están definidas para tomar valores numéricos, aun cuando describan eventos cualitativos. Por ejemplo, considere el lanzamiento de una moneda en el que los dos resultados son cara o cruz. Se puede definir una variable aleatoria como sigue: $X = 1$ si se obtiene cara y $X = 0$ si se obtiene cruz.

A una variable aleatoria que sólo puede tomar los valores 0 o 1 se le llama **variable aleatoria de Bernoulli** (o **binaria**). En la probabilidad básica, al evento $X = 1$ suele llamársele "éxito" y al evento $X = 0$ "fracaso". Para una aplicación particular, la nomenclatura éxito-fracaso puede que no corresponda a nuestra noción de éxito y fracaso, sin embargo, es una terminología útil que será adoptada en este libro.

Variables aleatorias discretas

Una **variable aleatoria discreta** es una variable que sólo toma una cantidad finita o una cantidad infinita contable de valores. La noción de "infinita contable" significa que aunque la variable aleatoria pueda tomar una cantidad infinita de valores, éstos se pueden poner en correspondencia uno a uno con los enteros positivos. Como la diferencia entre "contable infinito" e "incontable infinito" es un poco sutil, en este libro se concentrará la atención en variables aleatorias discretas que tomen sólo una cantidad finita de valores. Larsen y Marx (1986, capítulo 3) proporcionan un estudio detallado de este tema.

Una variable aleatoria de Bernoulli es el ejemplo más sencillo de variable aleatoria discreta. Lo único que se necesita para describir por completo el comportamiento de una variable aleatoria de Bernoulli es la probabilidad de que la variable tome el valor uno. En el ejemplo del lanzamiento de la moneda, si ésta es "legítima", entonces $P(X = 1) = 1/2$ (que se lee "la probabilidad de que X sea igual a uno es un medio"). Como las probabilidades deben sumar uno, también $P(X = 0) = 1/2$.

En las ciencias sociales, los científicos están interesados en algo más que el lanzamiento de una moneda, de manera que hay que considerar situaciones más generales. Considere, de nuevo, el ejemplo de la aerolínea que tiene que decidir cuántas reservaciones aceptar para un vuelo en el que hay 100 asientos disponibles. Este problema puede analizarse en el contexto de las variables aleatorias de Bernoulli de la manera siguiente: dado un cliente, tomado aleatoriamente, se define una variable aleatoria de Bernoulli como $X = 1$ si la persona utiliza su reservación y $X = 0$ si no la utiliza.

No hay razón para pensar que la probabilidad de que un determinado cliente utilice su reservación sea 1/2; en principio, esta probabilidad puede ser cualquier número entre cero y uno. Llamémosle θ a este número, de manera que

$$P(X = 1) = \theta \qquad \textbf{B.1}$$

$$P(X = 0) = 1 - \theta. \qquad \textbf{B.2}$$

Por ejemplo, si $\theta = .75$, entonces hay 75% de probabilidad de que el cliente utilice su reservación y 25% de que no la utilice. Intuitivamente se ve que el valor de θ es crucial para determinar la estrategia de reservaciones de la aerolínea. Los métodos para *estimar* θ, dados los datos históricos de las reservaciones en aerolíneas, son materia de la estadística matemática, tema que se verá en el apéndice C.

De manera más general, cualquier variable aleatoria discreta queda por completo descrita enumerando sus posibles valores así como las probabilidades de que tome cada uno de ellos. Si los valores X que puede tomar k son $\{x_1, ..., x_k\}$, entonces las probabilidades $p_1, p_2, ..., p_k$ se definen mediante

$$p_j = P(X = x_j),\ j = 1, 2, ..., k, \qquad \textbf{B.3}$$

donde cada p_j está entre 0 y 1 y

$$p_1 + p_2 + ... + p_k = 1. \qquad \textbf{B.4}$$

La ecuación (B.3) se lee: "la probabilidad de que X tome el valor x_j es igual a p_j."

Las ecuaciones (B.1) y (B.2) muestran que las probabilidades de éxito y fracaso de esta variable aleatoria de Bernoulli quedan totalmente determinadas por el valor de θ. Como las variables aleatorias de Bernoulli son tan frecuentes, se tiene una notación especial para ellas: $X \sim$ Bernoulli(θ) que se lee "X tiene una distribución de Bernoulli con probabilidad de éxito igual a θ."

The **función de densidad de probabilidad (fdp)** de X resume la información concerniente a los valores que puede tomar X y a sus correspondientes probabilidades:

$$f(x_j) = p_j,\ j = 1, 2, ..., k, \qquad \textbf{B.5}$$

con $f(x) = 0$ para toda x distinta de x_j para alguna j. En otras palabras, para cualquier número real x, $f(x)$ es la probabilidad de que la variable aleatoria X tome ese determinado valor. Cuando se tienen más de una variable para la fdp en cuestión se acostumbra usar subíndices: f_X es la fdp de X, f_Y es la fdp de Y, etcétera.

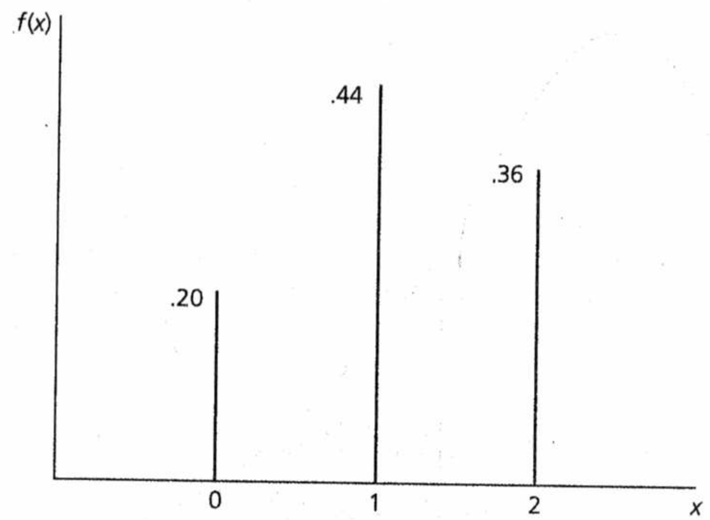
Dada la fdp de cualquier variable aleatoria discreta, es fácil calcular la probabilidad de cualquier evento relacionado con esa variable aleatoria. Por ejemplo, suponga que X es la cantidad de tiros libres anotados por un jugador de baloncesto en dos intentos, entonces X puede tomar los valores $\{0,1,2\}$. Suponga que la fdp de X está definida por

$$f(0) = .20,\ f(1) = .44 \text{ y } f(2) = .36.$$

Estas tres probabilidades suman uno, como debe ser. Usando esta fdp, la probabilidad de que el jugador anote *por lo menos* un tiro libre es: $P(X \geq 1) = P(X = 1) + P(X = 2) = .44 + .36 = .80$. En la figura B.1 se muestra la fdp de X.

FIGURA B.1

Fdp de la cantidad de tiros libres anotados en dos intentos.



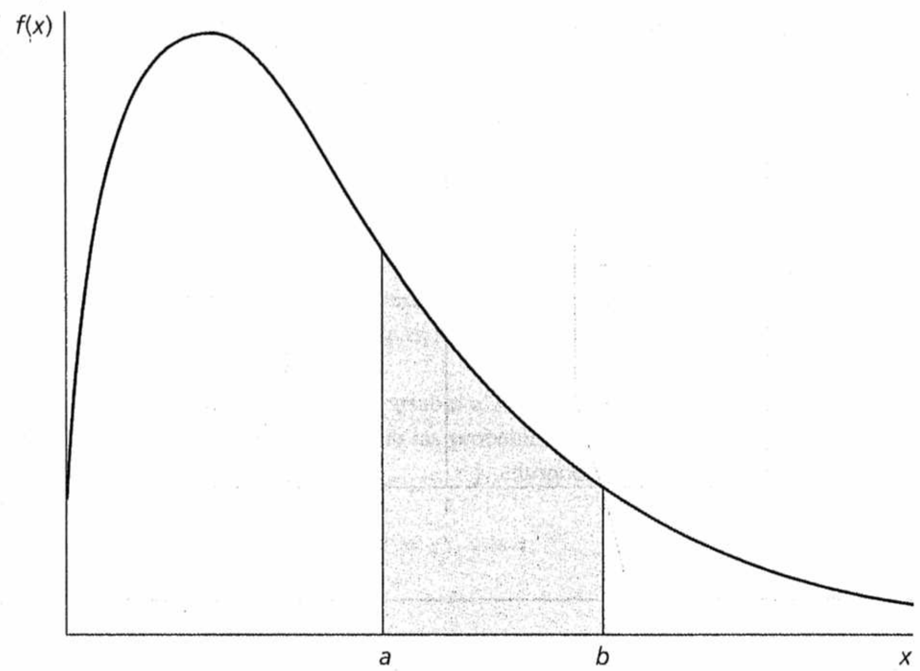
Variables aleatorias continuas

Una variable X es una **variable aleatoria continua** si la probabilidad de que la variable aleatoria tome cualquier valor real es *cero*. Esta definición es un poco contraintuitiva, ya que en cualquier aplicación se observará algún valor de la variable aleatoria. La idea es que los valores que puede tomar una variable aleatoria continua X son tantos que no es posible contarlos o hacerlos coincidir con los enteros positivos, de manera que la consistencia lógica indica que X puede tomar cada uno de estos valores con probabilidad cero. Aunque en la práctica las mediciones son siempre discretas, a las variables aleatorias que toman numerosos valores es mejor tratarlas como continuas. Por ejemplo, la medición más refinada del precio de un artículo se da en términos de centavos. Se puede pensar en enumerar por orden todos los valores de precio posibles (aun cuando esta lista quizá continue de forma indefinida), lo que técnicamente hace que el precio sea una variable aleatoria discreta. Sin embargo, son tantos los posibles valores de precio que usar la mecánica de las variables aleatorias discretas no es posible.

Para las variables aleatorias continuas también se puede definir una función de densidad de probabilidad y, como ocurre con las variables aleatorias discretas, la fdp proporciona información sobre los posibles valores de esta variable aleatoria. Sin embargo, dado que no tiene sentido analizar la posibilidad de que una variable aleatoria continua tome un determinado valor, la fdp de una variable de este tipo sólo se usa para calcular eventos que comprenden un rango de valores. Por ejemplo, si a y b son constantes y $a < b$, la probabilidad de que X se encuentre entre a y b, $P(a \leq X \leq b)$, es el *área* bajo la fdp entre los puntos a y b, como se muestra en la figura B.2. El lector familiarizado con el cálculo recordará que esta es la *integral* de la función f entre los puntos a y b. Toda el área bajo la fdp siempre debe ser igual a uno.

FIGURA B.2

Probabilidad de que *X* se encuentra entre los puntos *a* y *b*.



Para calcular probabilidades de variables aleatorias continuas, es más fácil emplear la **función de distribución acumulada (fda)**. Si X es cualquier variable aleatoria, entonces para cualquier número real x su fda está definida por

$$F(x) \equiv P(X \leq x). \qquad \textbf{B.6}$$

En el caso de variables aleatorias discretas, (B.6) se obtienen sumando las fdp de todos los valores x_j tales que $x_j \leq x$. En el caso de variables aleatorias continuas, $F(x)$ es el área bajo la fdp, f, a la izquierda del punto x. Como $F(x)$ es simplemente una probabilidad, su valor estará siempre entre 0 y 1. Además, si $x_1 < x_2$, entonces $P(X \leq x_1) \leq P(X \leq x_2)$, es decir $F(x_1) \leq F(x_2)$. Esto significa que una fda es una función creciente (o por lo menos no decreciente) de x.

Dos propiedades importantes de las fda útiles para calcular probabilidades son las siguientes:

$$\text{Para todo número } c,\ P(X > c) = 1 - F(c). \qquad \textbf{B.7}$$

$$\text{Para todo par de números } a < b,\ P(a < X \leq b) = F(b) - F(a). \qquad \textbf{B.8}$$

En nuestro estudio de la econometría, las fda sólo se usarán para calcular probabilidades de variables aleatorias continuas, en cuyo caso no importa si las desigualdades empleadas en las fórmulas de probabilidad son desigualdades estrictas o no. Es decir, dada una variable aleatoria continua X,

$$P(X \geq c) = P(X > c), \qquad \text{B.9}$$

y

$$P(a < X < b) = P(a \leq X \leq b) = P(a \leq X < b) = P(a < X \leq b). \qquad \text{B.10}$$

Las ecuaciones (B.9) y (B.10) combinadas con las ecuaciones (B.7) y (B.8) expanden enormemente los cálculos de probabilidad que pueden hacerse usando las fda.

Para todas las distribuciones continuas importantes en la probabilidad y en la estadística, las funciones de distribución acumulada ya han sido tabuladas. La más conocida de estas distribuciones es la normal, que se verá junto con otras en la sección B.5.

B.2 Distribuciones conjuntas, distribuciones condicionales e independencia

En economía, suele interesar la ocurrencia de eventos en los que participan más de una variable aleatoria. Por ejemplo, en el caso antes citado de las reservaciones de una aerolínea, a esta le puede interesar la probabilidad de que una persona que haga una reservación utilice su reservación *y* además sea un viajero de negocios; este es un ejemplo de *probabilidad conjunta*. O puede ocurrir que a la aerolínea le interese la *probabilidad condicional* siguiente: dada la condición de que la persona es viajero de negocios, ¿cuál es la probabilidad de que utilice su reservación? En las dos subsecciones siguientes, se formalizarán los conceptos de distribución conjunta y distribución condicional, así como el importante concepto de *independencia* de las variables aleatorias.

Distribuciones conjuntas e independencia

Sean X y Y variables aleatorias discretas. Entonces (X,Y) tiene una **distribución conjunta**, descrita completamente por la *función de densidad de probabilidad conjunta* de (X,Y):

$$f_{X,Y}(x,y) = P(X = x, Y = y), \qquad \text{B.11}$$

donde el lado derecho es la probabilidad de que $X = x$ y $Y = y$. Si X y Y son continuas, también se puede definir una fdp, pero estos detalles no se verán aquí, debido a que en este libro las fdp de variables aleatorias continuas no se emplean explícitamente.

Hay un caso en el que la fdp conjunta se obtiene con facilidad si se cuenta con las fdp de X y de Y. En particular, se dice que las variables aleatorias X y Y son independientes si y solo si,

$$f_{X,Y}(x,y) = f_X(x)f_Y(y) \qquad \text{B.12}$$

para toda x y y, donde f_X es la fdp de X y f_Y es la fdp de Y. Cuando se tiene más de una variable aleatoria, a las fdp f_X y f_Y se les suele llamar *funciones de densidad de probabilidad marginal* para distinguirlas de la fdp conjunta $f_{X,Y}$. Esta definición de independencia es válida tanto para variables discretas como para variables continuas.

Para entender el significado de (B.12), lo más fácil es ver el caso discreto. Si X y Y son discretas, entonces (B.12) es lo mismo que

$$P(X = x, Y = y) = P(X = x)\, P(Y = y); \qquad \text{B.13}$$

en otras palabras, la probabilidad de que $X = x$ y $Y = y$ es el producto de las probabilidades $P(X = x)$ y $P(Y = y)$. Una consecuencia de B.13 es que (cuando X y Y son independientes) las probabilidades conjuntas son bastante fáciles de calcular, ya que sólo es necesario conocer $P(X = x)$ y $P(Y = y)$.

Si las variables aleatorias no son independientes se dice que son *dependientes*.

Ejemplo B.1

[Lanzamiento de un tiro libre]

Imagine a un jugador de baloncesto que lanza dos tiros libres. Sea X una variable aleatoria de Bernoulli igual a uno si el jugador anota el primer tiro libre e igual a cero si no es así. Sea Y una variable aleatoria de Bernoulli igual a uno si el jugador anota el segundo tiro libre. Suponga que en tiros libres el jugador tiene una probabilidad de anotar de 80%, entonces $P(X = 1) = P(Y = 1) = .8$. ¿Cuál es la probabilidad de que el jugador anote los dos tiros libres?

Si X y Y son independientes, esta pregunta se responde fácilmente: $P(X = 1, Y = 1) = P(X = 1)\, P(Y = 1) = (.8)(.8) = .64$. Por tanto, la posibilidad de que anote los dos tiros libres es de 64%. Pero si la probabilidad de anotar el segundo tiro libre depende de si anota o no el primer tiro libre —es decir, si X y Y no son independientes— entonces este sencillo cálculo no es válido.

La independencia de variables aleatorias es un concepto muy importante. En la subsección siguiente se mostrará que si X y Y son independientes, conocer el valor de X no modifica las probabilidades de los valores posibles de Y, y viceversa. Un hecho útil acerca de la independencia es que si X y Y son independientes, y para funciones g y h cualesquiera se definen las nuevas variables aleatorias $g(X)$ y $h(Y)$, entonces las nuevas variables aleatorias también son independientes.

Pero no es necesario quedarse con dos variables. Si $X_1, X_2, \ldots, X_n$ son variables aleatorias discretas, entonces su fdp conjunta es $f(x_1, x_2, \ldots, x_n) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$. Las variables aleatorias $X_1, X_2, \ldots, X_n$ son **variables aleatorias independientes** si y solo si su fdp conjunta es el producto de sus fdp individuales para toda $(x_1, x_2, \ldots, x_n)$. Esta definición de independencia también es válida para variables aleatorias continuas.

El concepto de independencia es muy importante en la probabilidad y la estadística en la obtención de algunas de las distribuciones clásicas. Ya antes se definió una variable aleatoria de Bernoulli como una variable aleatoria uno-cero que indica la ocurrencia o no de un evento. Con frecuencia, lo que interesa es la cantidad de éxitos en una sucesión de ensayos *independientes* de Bernoulli. El ejemplo más usual de estos ensayos es el lanzamiento de una moneda, una y otra vez. Como el resultado de cada lanzamiento no tiene nada que ver con los resultados de los otros lanzamientos, la independencia es un supuesto adecuado.

En situaciones más complicadas, la independencia suele ser una aproximación razonable. En el ejemplo de las reservaciones de la aerolínea, suponga que para un determinado vuelo la aerolínea acepta n reservaciones. Para $i = 1, 2, \ldots, n$, sea Y_i la variable aleatoria de Bernoulli que indica si el cliente i utiliza su reservación o no: $Y_i = 1$ si el cliente i utiliza su reservación y $Y_i = 0$ si no la utiliza. Sea θ la probabilidad de éxito (el cliente utiliza su reservación); para cada Y_i hay una distribución de Bernoulli(θ). Como aproximación, puede suponerse que las Y_i son independientes unas de otras, aunque en realidad puede que esto no sea verdad: algunas personas viajan en grupo, lo que significa que, el que una persona utilice o no su reservación no es completamente independiente de si las otras la utilizan. Como modelar este tipo de dependencia es complicado, se puede aceptar usar la independencia como aproximación.

La variable que interesa es la cantidad total de los n clientes que utiliza su reservación; llámese X a esta variable. Como cada Y_i es una unidad en el momento que la persona utiliza o no su reservación, se puede escribir $X = Y_1 + Y_2 + \ldots + Y_n$. Ahora, suponiendo que la probabilidad de éxito de cada Y_i sea θ y que las Y_i sean independientes, se pueden mostrar que X tiene una **distribución binomial**. Es decir, la función de densidad de probabilidad de X es

$$f(x) = \binom{n}{x}\theta^x(1-\theta)^{n-x},\ x = 0, 1, 2, \ldots, n, \tag{B.14}$$

donde $\binom{n}{x} = \frac{n!}{x!(n-x)!}$, y para todo entero n, $n!$ (que se lee "n factorial") está definido como $n! = n\cdot(n-1)\cdot(n-2)\cdots 1$. Por definición, $0! = 1$. Si una variable aleatoria X tiene la fdp dada por (B.14), se escribe $X \sim$ Binomial (n, θ). La ecuación (B.14) se emplea para calcular $P(X = x)$ para cualquier valor de x desde 0 hasta n.

Si el vuelo dispone de 100 asientos, a la aerolínea le interesa $P(X > 100)$. Suponga que inicialmente $n = 120$, es decir, que la aerolínea ha aceptado 120 reservaciones, y que la probabilidad de que cada persona utilice su reservación es $\theta = .85$. Entonces, $P(X > 100) = P(X = 101) + P(X = 102) + \ldots + P(X = 120)$ y cada una de las probabilidades de esta suma se encuentra empleando la ecuación (B.14) con $n = 120$, $\theta = .85$ y el valor de x correspondiente (desde 101 hasta 120). Hacer estos cálculos a mano es bastante complicado, pero por fortuna muchos paquetes de estadística tienen comandos para calcular este tipo de probabilidades. En este caso, la probabilidad de que más de 100 personas utilicen su reservación es aproximadamente .659, que es tal vez un riesgo de exceso de reservaciones mayor al que la aerolínea estaría dispuesta a tolerar. En cambio, si la cantidad de reservaciones es 110, la probabilidad de que más de 100 pasajeros utilice su reservación es aproximadamente de sólo .024.

Distribuciones condicionales

En econometría, usualmente interesa saber cómo está relacionada una variable, a la que se le llamará Y, con otra u otras variables. Por ahora, suponga que sólo interesa el efecto de una variable, llámesele X. Lo más que se puede saber acerca de cómo afecta X a Y está contenido en la **distribución condicional** de Y dada X. Esta información está resumida en la *función de densidad de probabilidad condicional*, definida por

$$f_{Y|X}(y|x) = f_{X,Y}(x,y)/f_X(x) \tag{B.15}$$

para todos los valores de x tales que $f_X(x) > 0$. La interpretación de la ecuación (B.15) se ve más fácilmente cuando X y Y son discretas. Entonces,

$$f_{Y|X}(y|x) = P(Y = y|X = x), \tag{B.16}$$

donde el lado derecho se lee como "la probabilidad de que $Y = y$ dado que $X = x$". Cuando Y es continua, $f_{Y|X}(y|x)$ no se puede interpretar directamente como una probabilidad, por las razones antes vistas, sino que las probabilidades condicionales se encuentran calculando áreas bajo la fdp condicional.

Una característica importante de las distribuciones condicionales es que, si X y Y son variables aleatorias independientes, conocer el valor que toma X no dice nada acerca de la probabilidad de que Y tome diversos valores (y viceversa). Es decir, $f_{Y|X}(y|x) = f_Y(y)$ y $f_{X|Y}(x|y) = f_X(x)$.

Ejemplo B.2

[Lanzamiento de un tiro libre]

Considere nuevamente el ejemplo del basquetbolista, en el que se tienen dos tiros libres. Suponga que la densidad condicional es

$$f_{Y|X}(1|1) = .85, f_{Y|X}(0|1) = .15$$
$$f_{Y|X}(1|0) = .70, f_{Y|X}(0|0) = .30.$$

Esto significa que la probabilidad de que un jugador anote el segundo tiro libre depende de si anotó o no el primero: si anotó éste, la posibilidad de que anote el segundo es .85; si lo falló, la posibilidad de que anote el segundo es .70. Esto implica que X y Y *no* son independientes; son dependientes.

Se puede calcular $P(X = 1, Y = 1)$ siempre y cuando se conozca $P(X = 1)$. Suponga que la probabilidad de anotar el primer tiro libre es .8, es decir, $P(X = 1) = .8$. Entonces, de acuerdo con la ecuación (B.15), se tiene

$$P(X = 1, Y = 1) = P(Y = 1|X = 1) \cdot P(X = 1) = (.85)(.8) = .68.$$

B.3 Características de las distribuciones de probabilidad

En muchos casos, únicamente interesan aquí algunas características de las distribuciones de las variables aleatorias. Estas características de interés se pueden agrupar en tres categorías: medidas de tendencia central, medidas de variabilidad o dispersión y medidas de la relación entre dos variables aleatorias. La última de estas categorías se verá en la sección B.4.

Una medida de tendencia central: el valor esperado

El valor esperado es uno de los conceptos más importantes de la probabilidad que surgen en el estudio de la econometría. Si X es una variable aleatoria, el **valor esperado** (o la esperanza) de X, que se denota $E(X)$ o, alguna veces, μX o simplemente μ, es un promedio ponderado de todos los posibles valores de X. Los pesos de ponderación están determinados por la función de densidad de probabilidad. Al valor esperado también se le suele llamar *media poblacional*, en especial cuando se quiere hacer énfasis en que X representa una variable poblacional.

La definición precisa de valor esperado es más sencilla de dar cuando X es una variable aleatoria discreta que toma un número finito de valores, por ejemplo, $\{x_1, \ldots, x_k\}$. Sea $f(x)$ la función de densidad de probabilidad de X. El valor esperado de X es el promedio ponderado

$$E(X) = x_1 f(x_1) + x_2 f(x_2) + \ldots + x_k f(x_k) \equiv \sum_{j=1}^{k} x_j f(x_j). \qquad \boxed{\textbf{B.17}}$$

Este valor es fácil de calcular cuando se conocen todos los valores de la fdp para cada uno de los valores que puede tomar X.

Ejemplo B.3

[Cálculo del valor esperado]

Suponga que X toma los valores -1, 0 y 2 con probabilidades 1/8, 1/2 y 3/8, respectivamente. Entonces

$$E(X) = (-1)\cdot(1/8) + 0\cdot(1/2) + 2\cdot(3/8) = 5/8.$$

Este ejemplo ilustra algo curioso acerca del valor esperado: el valor esperado de X puede ser un número que no sea ninguno de los posibles valores de X. Se sabe que X toma los valores -1, 0 y 2, y sin embargo, su valor esperado es 5/8. Esto hace al valor esperado inadecuado para resumir la tendencia central de ciertas variables aleatorias discretas; sin embargo cálculos como los recién mencionados suelen ser útiles, como se verá más adelante.

Si X es una variable aleatoria continua, entonces $E(X)$ está definida como una integral:

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx, \qquad \textbf{B.18}$$

que se supone está bien definida. Esta integral también se interpreta como un promedio ponderado. En la mayoría de las distribuciones continuas más comunes, $E(X)$ es un número cuyo valor es un valor posible de X. En este libro no será necesario calcular valores esperados usando la integración, pero sí se recurrirá a algunas de las fórmulas más conocidas de la probabilidad para los valores esperados de variables aleatorias especiales.

Dada una variable aleatoria X y una función $g(\cdot)$, se puede crear una nueva variable aleatoria $g(X)$. Por ejemplo, si X es una variable aleatoria, entonces también X^2 y $\log(X)$ (si $X > 0$) son variables aleatorias. El valor esperado de $g(X)$ es, otra vez, un simple promedio ponderado:

$$E[g(X)] = \sum_{j=1}^{k} g(x_j)f_X(x_j) \qquad \textbf{B.19}$$

o, para variables aleatorias continuas,

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f_X(x)dx. \qquad \textbf{B.20}$$

Ejemplo B.4

[Valor esperado de X^2]

En el caso de la variable aleatoria del ejemplo B.3, sea $g(X) = X^2$. Entonces,

$$E(X^2) = (-1)^2(1/8) + (0)^2(1/2) + (2)^2(3/8) = 13/8.$$

En el ejemplo B.3, se obtuvo $E(X) = 5/8$, de manera que $[E(X)]^2 = 25/64$. Esto demuestra que $E(X^2)$ *no* es igual a $[E(X)]^2$. Efectivamente, si $g(X)$ no es una función lineal $E[g(X)] \neq g[E(X)]$ (salvo en casos muy especiales).

Si X y Y son variables aleatorias, entonces para toda función g $g(X,Y)$ es una variable aleatoria y por tanto, se puede definir su esperanza. Si tanto X como Y son discretas, y toman los valores $\{x_1, x_2, \ldots, x_k\}$ y $\{y_1, y_2, \ldots, y_m\}$, respectivamente, su valor esperado es

$$E[g(X,Y)] = \sum_{h=1}^{k}\sum_{j=1}^{m} g(x_h,y_j)f_{X,Y}(x_h,y_j),$$

donde $f_{X,Y}$ es la fdp de (X,Y). En el caso de variables aleatorias continuas, la definición es más complicada, ya que se requiere la integración; pero no la necesitaremos aquí. La extensión a más de dos variables aleatorias es inmediata.

Propiedades de los valores esperados

En econometría, lo que más interesa no es el cálculo de los valores esperados de las distintas distribuciones; los principales cálculos ya se han hecho muchas veces y se puede confiar en ellos. Lo que se necesita es manipular algunos valores esperados empleando algunas reglas sencillas. Estas reglas son tan importantes que aquí se les han colocado etiquetas:

Propiedad E.1: Para toda constante c, $\mathrm{E}(c) = c$.

Propiedad E.2: Para todo par de constantes a y b, $\mathrm{E}(aX + b) = a\mathrm{E}(X) + b$.

Una consecuencia importantes de E.2 es que, si $\mu = \mathrm{E}(X)$, y se define una nueva variable aleatoria $Y = X - \mu$, entonces $\mathrm{E}(Y) = 0$; en E.2 hágase $a = 1$ y $b = -\mu$.

Como ejemplo de la propiedad E.2, sea X la temperatura medida en grados Celsius hacia la tarde de un día y en un lugar determinado; suponga que la temperatura esperada es $\mathrm{E}(X) = 25$. Si Y es la temperatura medida en grados Fahrenheit, entonces $Y = 32 + (9/5)X$. De acuerdo con la propiedad E.2, la temperatura esperada en grados Farenheit es $\mathrm{E}(Y) = 32 + (9/5)\cdot\mathrm{E}(X) = 32 + (9/5)\cdot 25 = 77$.

Por lo general, es fácil calcular el valor esperado de una función lineal de muchas variables aleatorias.

Propiedad E.3: Si $\{a_1, a_2, \ldots, a_n\}$ son constantes y $\{X_1, X_2, \ldots, X_n\}$ son variables aleatorias, entonces

$$\mathrm{E}(a_1X_1 + a_2X_2 + \ldots + a_nX_n) = a_1\mathrm{E}(X_1) + a_2\mathrm{E}(X_2) + \ldots + a_n\mathrm{E}(X_n).$$

O, empleando la sumatoria,

$$\mathrm{E}\left(\sum_{i=1}^{n} a_iX_i\right) = \sum_{i=1}^{n} a_i\mathrm{E}(X_i). \qquad \boxed{\textbf{B.21}}$$

Como caso especial de esto, se tiene (con todas las $a_i = 1$)

$$\mathrm{E}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{E}(X_i). \qquad \boxed{\textbf{B.22}}$$

de manera que el valor esperado de una suma es la suma de los valores esperados. Esta propiedad es usada con frecuencia para derivaciones en estadística matemática.

Ejemplo B.5

[Determinar ingresos esperados]

Sean X_1, X_2 y X_3 las cantidades de pizzas pequeñas, medianas y grandes vendidas en una pizzería en un día determinado. Éstas son variables aleatorias cuyos valores esperados son $\mathrm{E}(X_1) = 25$, $\mathrm{E}(X_2) = 57$ y

$E(X_3) = 40$. Los precios de las pizzas pequeñas, medianas y grandes son \$5.50, \$7.60 y \$9.15 respectivamente. Por tanto, el ingreso esperado por día de la venta de las pizzas es

$$\begin{aligned} E(5.50\,X_1 + 7.60\,X_2 + 9.15\,X_3) &= 5.50\,E(X_1) + 7.60\,E(X_2) + 9.15\,E(X_3) \\ &= 5.50(25) + 7.60(57) + 9.15(40) = 936.70, \end{aligned}$$

es decir, \$936.70. El ingreso real en un día determinado será, por lo general, distinto de este valor, pero este es el valor *esperado*.

La propiedad E.3 también puede usarse para mostrar que si $X \sim$ Binomial (n,θ), entonces $E(X) = n\theta$. Es decir, la cantidad esperada de éxitos en n ensayos de Bernoulli es simplemente la cantidad de ensayos multiplicada por la probabilidad de éxito en un ensayo particular. Esto se puede ver con facilidad escribiendo X como $X = Y_1 + Y_2 + \ldots + Y_n$, donde cada $Y_i \sim$ binomial(i,n). Entonces,

$$E(X) = \sum_{i=1}^{n} E(Y_i) = \sum_{i=1}^{n} \theta = n\theta.$$

Esto se puede aplicar al ejemplo de las reservaciones de la aerolínea, donde ésta hace $n = 120$ reservaciones y la probabilidad de que el cliente utilice su reservación es $\theta = .85$. La cantidad *esperada* de personas que utilizará su reservación es $120\,(.85) = 102$. Por lo tanto, si se dispone de 100 asientos, la cantidad esperada de personas que usarán su reservación es demasiada; esto puede servir de orientación a la aerolínea para decidir si es buena idea hacer 120 reservaciones.

En realidad, lo que la aerolínea deberá hacer es definir una función de utilidad que tome en cuenta el ingreso neto por asiento vendido y el costo por pasajero al que se le niegue el embarque. Esta función de utilidad es aleatoria debido a que la cantidad de personas que realmente utilicen su reservación es aleatoria. Sea r el ingreso neto por pasajero. (Para simplificar este se puede considerar como el precio del boleto.) Sea c la compensación que se le debe a cada pasajero al que se le niega el embarque. Ni r ni c son aleatorios; se supone que la aerolínea conoce estas cantidades. Sea Y la ganancia del vuelo. Entonces, como se dispone de 100 asientos,

$$\begin{aligned} Y &= rX \text{ si } X \leq 100 \\ &= 100r - c(X - 100) \text{ si } X > 100. \end{aligned}$$

La primera ecuación da la ganancia si no se presentan más de 100 personas para el vuelo; la segunda ecuación da la ganancia si más de 100 personas quieren utilizar su reservación. (En el último caso, la ganancia neta de la venta de los boletos es $100r$, ya que se han vendido los 100 boletos y $c(X - 100)$ es el costo de hacer más de 100 reservaciones.) Empleando el hecho de que X tiene una distribución Binomial$(n, .85)$ donde n es la cantidad de reservaciones hecha, la ganancia esperada, $E(Y)$, puede determinarse como función de n (y de r y c). Calcular $E(Y)$ de forma directa quizá resulte difícil, pero empleando una computadora esta cantidad puede determinarse rápidamente. Una vez dados los valores de r y c, el valor de n que maximiza la ganancia esperada puede calcularse buscando entre los diferentes valores de n.

Otra medida de tendencia central: la mediana

El valor esperado es sólo una de las posibilidades para definir la tendencia central de una variable aleatoria. Otra medida de tendencia central es la **mediana**. Una definición general de la *mediana* es demasiado complicada para los propósitos de este libro. Si X es continua, entonces la mediana de X, llámese m, es el valor tal que una mitad del área bajo la curva de la fdp queda a la izquierda de m y la otra mitad del área queda a la derecha de m.

Si X es discreta y toma una cantidad finita de valores, la mediana se obtiene ordenando todos los posibles valores de X y seleccionando después el valor de en medio. Por ejemplo, si

X toma los valores $\{-4, 0, 2, 8, 10, 13, 17\}$, entonces la mediana de X es 8. Pero si X toma un número par de valores, hay en realidad dos valores de en medio, estos suelen promediarse para obtener un único valor para la mediana. Por tanto, si X toma los valores $\{-5,3,9,17\}$, entonces los valores medianos son 3 y 9; si se determina el promedio de estos dos valores se obtiene un valor mediano que es igual a 6.

En general, la mediana, que suele denotarse Med(X), y el valor esperado, E(X), suelen ser diferentes. Ninguno de los dos es "mejor" que el otro como medida de tendencia central; ambos son adecuados como medida del centro de la distribución de X. Hay un caso especial, en el que la mediana y el valor esperado (o media) son iguales. Si X tiene una **distribución simétrica** respecto al valor μ, entonces μ es tanto el valor esperado como la mediana. Matemáticamente, la condición es $f(\mu + x) = f(\mu - x)$ para toda x. En la figura B.3 se ilustra este caso.

Medidas de variabilidad: varianza y desviación estándar

Aunque la tendencia central de una variable aleatoria es valiosa, no da toda la información deseada acerca de la distribución de la variable aleatoria. En la figura B.4 se muestran las fdp de dos variables aleatorias que tienen la misma media. Se ve con claridad que la distribución de X está más estrechamente centrada en torno a la media que la distribución de Y. Se desea tener una manera sencilla de resumir las diferencias en la dispersión de las distribuciones.

Varianza

Dada una variable aleatoria X, sea $\mu = \mathrm{E}(X)$. Hay varias maneras de medir qué tan lejos está X de su valor esperado, pero la forma más sencilla para manipular algebraicamente es el cuadrado

FIGURA B.3

Una distribución de probabilidad simétrica.

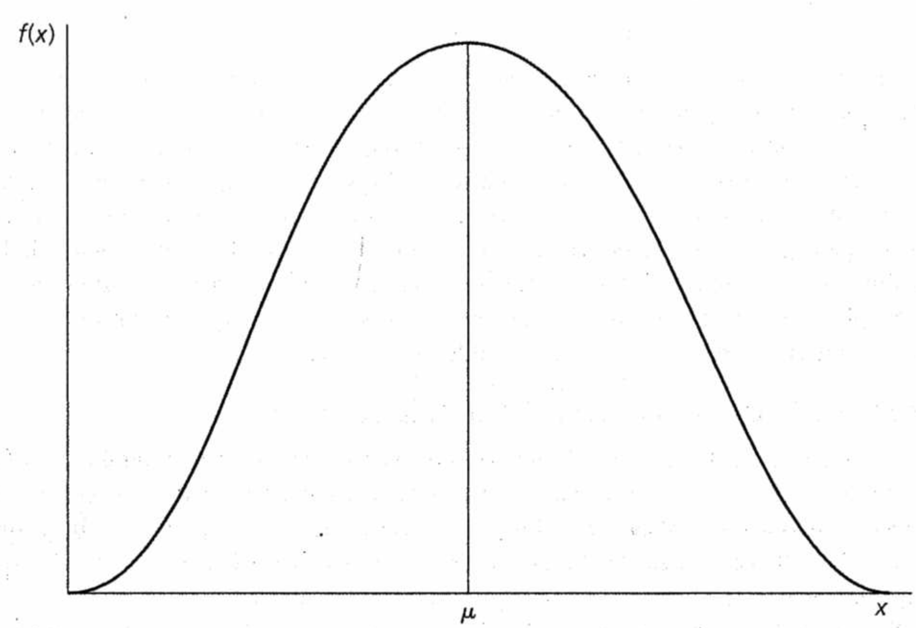
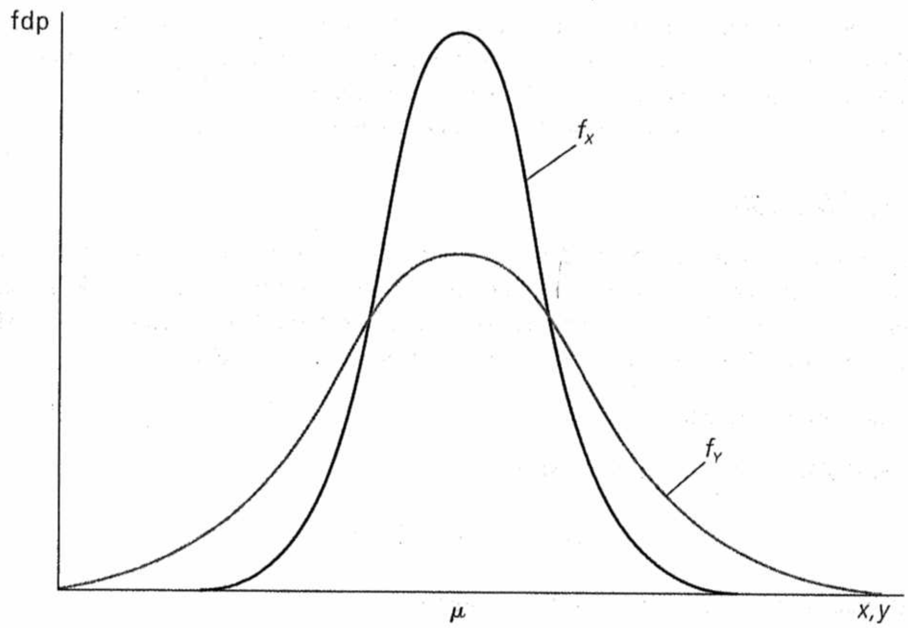


FIGURA B.4

Variables aleatorias con la misma media pero con distintas distribuciones.



de la diferencia, $(X - \mu)^2$. (Al elevar al cuadrado se elimina el signo de la distancia medida; el valor positivo que se obtiene corresponde a la noción intuitiva de distancia y trata de manera simétrica a valores mayores y menores que μ.) Esta distancia también es una variable aleatoria, pues cambia con cada valor que toma X. Así como se necesita un número que resuma la tendencia central de X, también se requiere uno que indique qué tan lejos, *en promedio*, está X de μ. Este número es la **varianza**, que da la distancia esperada de X a su media:

$$\mathrm{Var}(X) \equiv \mathrm{E}[(X - \mu)^2]. \qquad \textbf{B.23}$$

La varianza también suele denotarse como σ_X^2, o simplemente σ^2, cuando el contexto es claro. De acuerdo con la ecuación (B.23) se tiene que la varianza siempre es no negativa.

Es útil observar que

$$\sigma^2 = \mathrm{E}(X^2 - 2X\mu + \mu^2) = \mathrm{E}(X^2) - 2\mu^2 + \mu^2 = \mathrm{E}(X^2) - \mu^2. \qquad \textbf{B.24}$$

Cuando se usa, ya sea (B.23) o (B.24), no es necesario distinguir entre variables aleatorias discretas y variables aleatorias continuas; la definición de varianza es la misma en ambos casos. Lo más frecuente es que se calcule $\mathrm{E}(X)$, después $\mathrm{E}(X^2)$, y después se utiliza la fórmula (B.24). Por ejemplo, si $X \sim \mathrm{Bernoulli}(\theta)$, entonces $\mathrm{E}(X) = \theta$, y, como $X^2 = X$, $\mathrm{E}(X^2) = \theta$. Por tanto, de acuerdo con la ecuación (B.24 $\mathrm{Var}(X) = \mathrm{E}(X^2) - \mu^2 = \theta - \theta^2 = \theta(1 - \theta)$.

A continuación se presentan dos propiedades importantes de la varianza.

Propiedad VAR.1: $\mathrm{Var}(X) = 0$ si y solo si, existe una constante c, tal que $\mathrm{P}(X = c) = 1$, en cuyo caso, $\mathrm{E}(X) = c$.

Esta propiedad indica que la varianza de cualquier constante es cero y que si una variable aleatoria tiene varianza cero, entonces es una constante.

Propiedad VAR.2: Para cualesquiera constantes a y b, $\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$.

Esto significa que sumar una constante a una variable aleatoria no modifica la varianza, pero multiplicar una variable aleatoria por una constante aumenta la varianza en un factor igual al *cuadrado* de la constante. Por ejemplo, si X *denota temperatura en grados Celsius y* $Y = 32 + (9/5)X$ es temperatura en grados Fahrenheit, entonces $\mathrm{Var}(Y) = (9/5)^2\mathrm{Var}(X) = (81/25)\mathrm{Var}(X)$.

Desviación estándar

La **desviación estándar** de una variable aleatoria, que se denota $\mathrm{sd}(X)$, es la raíz cuadrada positiva de la varianza: $\mathrm{sd}(X) \equiv +\sqrt{\mathrm{Var}(X)}$. La desviación estándar también se suele denotar σ_X, o simplemente σ, cuando se sobreentiende la variable aleatoria. De las propiedades VAR.1 y VAR.2 se deducen de inmediato dos propiedades de la desviación estándar.

Propiedad SD.1: Para toda constante c, $\mathrm{sd}(c) = 0$.

Propiedad SD.2: Para todas las constantes a y b,

$$\mathrm{sd}(aX + b) = |a|\mathrm{sd}(X).$$

En particular, si $a > 0$, entonces $\mathrm{sd}(aX) = a \cdot \mathrm{sd}(X)$.

Esta última propiedad hace que resulte más natural trabajar con la desviación estándar que con la varianza. Por ejemplo, suponga que X es una variable aleatoria medida en miles de dólares, como por ejemplo, ingreso. Si se define $Y = 1{,}000X$, entonces Y es ingreso medido en dólares. Suponga que $\mathrm{E}(X) = 20$ y que $\mathrm{sd}(X) = 6$. Entonces $\mathrm{E}(Y) = 1{,}000\mathrm{E}(X) = 20{,}000$ y $\mathrm{sd}(Y) = 1{,}000 \cdot \mathrm{sd}(X) = 6{,}000$, de manera que tanto el valor esperado como la desviación estándar aumentan en un mismo factor, 1,000. Si se calcula la varianza, se tiene $\mathrm{Var}(Y) = (1{,}000)^2\mathrm{Var}(X)$, de manera que la varianza de Y es un millón de veces más grande que la de X.

Estandarización de una variable aleatoria

Como aplicación de las propiedades de la varianza y de la desviación estándar –y un tema de interés práctico por sí mismo– suponga que dada una variable aleatoria X, se define una nueva variable aleatoria restándole su media μ y dividiendo entre su desviación estándar σ:

$$Z \equiv \frac{X - \mu}{\sigma}, \qquad \boxed{\text{B.25}}$$

lo cual puede escribirse como $Z = aX + b$, donde $a \equiv (1/\sigma)$ y $b \equiv -(\mu/\sigma)$. Entonces, de acuerdo con la propiedad E.2,

$$\mathrm{E}(Z) = a\mathrm{E}(X) + b = (\mu/\sigma) - (\mu/\sigma) = 0.$$

De acuerdo con la propiedad VAR.2,

$$\mathrm{Var}(Z) = a^2\mathrm{Var}(X) = (\sigma^2/\sigma^2) = 1.$$

Por tanto, la variable aleatoria Z tiene media cero y varianza (y por tanto desviación estándar) igual a uno. A este proceso suele conocérsele como *estandarización* de la variable aleatoria X y a Z como **variable aleatoria estandarizada**. (En cursos de introducción a la estadística a veces es llamada *transformada-z* de X.) Es importante recordar que en el denominador de la ecuación (B.25) aparece la desviación estándar, no la varianza. Como se verá, esta transformación se usa con frecuencia en la inferencia estadística.

Como ejemplo concreto, considere que $\mathrm{E}(X) = 2$ y $\mathrm{Var}(X) = 9$. Entonces, el valor esperado de $Z = (X - 2)/3$ es cero y su varianza es uno.

Sesgo y curtosis

La versión estandarizada de una variable aleatoria puede utilizarse para definir otras características de la distribución de una variable aleatoria. Estas características se describen usando los llamados *momentos de orden superior*. Por ejemplo, el tercer momento de la variable aleatoria Z de la ecuación (B.25) se usa para determinar si la distribución es simétrica respecto a su media. Se puede escribir

$$\mathrm{E}(Z^3) = \mathrm{E}[(X - \mu)^3]/\sigma^3$$

Si X tiene una distribución simétrica respecto a μ, entonces Z tiene una distribución simétrica respecto a cero. (La división entre σ^3 no cambia si la distribución es simétrica o no.) Esto quiere decir que la densidad de Z en dos puntos z y $-z$ es la misma, lo que significa que, al calcular $\mathrm{E}(Z^3)$, los valores positivos de z^3 cuando $z > 0$ se compensan exactamente con los valores negativos $(-z)^3 = -z^3$. Por tanto, si X es simétrica respecto a cero, entonces $\mathrm{E}(Z) = 0$. Por lo general, $\mathrm{E}[(X - \mu)^3]/\sigma^3$ es considerada una medida del **sesgo** de la distribución de X. En la estadística se suele estimar $\mathrm{E}(Z^3)$ para determinar si la distribución de la población subyacente parece ser simétrica. (El ejercicio para computadora C5.4 del capítulo 5 proporciona un ejemplo.)

El cuarto momento de Z también puede ser informativo,

$$\mathrm{E}(Z^4) = \mathrm{E}[(X - \mu)^4]/\sigma^4.$$

Como $Z^4 \geq 0$, $\mathrm{E}(Z^4) \geq 0$ (y, en los casos interesantes, estrictamente mayor que cero). Sin un valor de referencia, es difícil interpretar los valores de $\mathrm{E}(Z^4)$, pero valores mayores significan que las colas de la distribución de X son más gruesas. El cuarto momento $\mathrm{E}(Z^4)$ se considera una medida de la **curtosis** de la distribución de X. En la sección B.5 se obtendrá el valor de $\mathrm{E}(Z^4)$ para una distribución normal.

B.4 Características de las distribuciones conjuntas y de las condicionales

Medidas de asociación: covarianza y correlación

Aunque la fdp conjunta de dos variables aleatorias describe completamente la relación entre ellas, es útil tener una medida resumida del promedio de la variación de estas dos variables aleatorias, una respecto a la otra. Como en el caso del valor esperado y de la varianza, éste es un solo número que resume algo acerca de toda una población, la que, en este caso, es la distribución conjunta de dos variables aleatorias.

Covarianza

Sean $\mu_X = \mathrm{E}(X)$ y $\mu_Y = \mathrm{E}(Y)$ y considere la variable aleatoria $(X - \mu_X)(Y - \mu_Y)$. Si X es mayor a su media y Y es mayor a su media, entonces $(X - \mu_X)(Y - \mu_Y) > 0$. Esto también es cierto si

$X < \mu_X$ y $Y < \mu_Y$. Por otro lado, si $X > \mu_X$ y $Y < \mu_Y$, o viceversa, entonces $(X - \mu_X)(Y - \mu_Y) < 0$. ¿Cómo puede indicar este producto algo acerca de la relación entre X y Y?

La **covarianza** entre dos variables aleatorias X y Y, llamada algunas veces *covarianza poblacional* para hacer énfasis en que se refiere a la relación entre dos variables que describen una población, está definida como el valor esperado del producto $(X - \mu_X)(Y - \mu_Y)$:

$$\mathrm{Cov}(X,Y) \equiv \mathrm{E}[(X - \mu_X)(Y - \mu_Y)], \tag{B.26}$$

que también suele denotarse σ_{XY}. Si $\sigma_{XY} > 0$, entonces, en promedio, cuando X es mayor a su media, también Y es mayor a su media. Si $\sigma_{XY} < 0$, entonces, en promedio, cuando X es mayor a su media Y es menor a su media.

Algunas expresiones útiles para calcular $\mathrm{Cov}(X,Y)$ son las siguientes:

$$\begin{aligned}\mathrm{Cov}(X,Y) &= \mathrm{E}[(X - \mu_X)(Y - \mu_Y)] = \mathrm{E}[(X - \mu_X)Y] \\ &= \mathrm{E}[X(Y - \mu_Y)] = \mathrm{E}(XY) - \mu_X\mu_Y.\end{aligned} \tag{B.27}$$

De (B.27) se sigue que si $\mathrm{E}(X) = 0$ o $\mathrm{E}(Y) = 0$, entonces $\mathrm{Cov}(X, Y) = \mathrm{E}(XY)$.

La covarianza es una medida de la dependencia *lineal* entre dos variables aleatorias. Si la covarianza es positiva indica que las dos variables aleatorias se mueven en la misma dirección, mientras que si es negativa indica que las dos variables se mueven en direcciones opuestas. Interpretar la *magnitud* de la covarianza puede resultar un poco difícil, como se verá en breve.

Como la covarianza es una medida de cómo están relacionadas dos variables, es natural preguntar por la relación entre la covarianza y el concepto de independencia. Esta relación está dada por la siguiente propiedad.

Propiedad COV.1: Si X y Y son independientes, entonces $\mathrm{Cov}(X,Y) = 0$.

Esta propiedad es consecuencia de la ecuación (B.27) y del hecho de que cuando X y Y son independientes, entondes $\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$. Es importante recordar que el inverso de COV.1 *no* es verdad: cero covarianza entre X y Y no implica que X y Y sean independientes. En efecto, hay variables aleatorias X tales que si $Y = X^2$, $\mathrm{Cov}(X,Y) = 0$. [Toda variable aleatoria para la que $\mathrm{E}(X) = 0$ y $\mathrm{E}(X^3) = 0$ tiene esta propiedad.] Si $Y = X^2$, entonces es claro que X y Y no son independientes: una vez que se conoce X, se conoce Y. Resulta bastante raro que X y X^2 tengan covarianza cero, y esto revela una debilidad de la covarianza como medida general de la asociación entre dos variables aleatorias. La covarianza es útil en contextos en los que las relaciones son por lo menos aproximadamente lineales.

La segunda propiedad importante de la covarianza está relacionada con covarianzas entre funciones lineales.

Propiedad COV.2: Para todas las constantes a_1, b_1, a_2 y b_2,

$$\mathrm{Cov}(a_1X + b_1, a_2Y + b_2) = a_1a_2\mathrm{Cov}(X, Y). \tag{B.28}$$

Una consecuencia importante de cov.2 es que la covarianza entre dos variables aleatorias puede alterarse simplemente multiplicando una o las dos variables aleatorias por una constante. Esto es importante en economía debido a que las variables monetarias, las tasas de inflación, etcétera, pueden ser definidas en diversas unidades de medición sin que esto cambie su significado.

Por último, es útil saber que el valor absoluto de la covarianza entre dos variables aleatorias está acotado por el producto de sus desviaciones estándar; esto se conoce como *desigualdad de Cauchy-Schwart*.

Propiedad COV.3: $|\text{Cov}(X,Y)| \leq \text{de}(X)\ \text{de}(Y)$.

Coeficiente de correlación

Suponga que se quiere conocer la relación entre cantidad de educación e ingreso anual en la población trabajadora. Para esto, educación se denota por X, e ingreso por Y, y se calcula su covarianza. Pero la respuesta que se obtenga dependerá de cómo se midan educación e ingreso. La propiedad COV.2 implica que la covarianza entre educación e ingresos depende de si éste se mide en dólares o en miles de dólares y de si la educación se mide en meses o en años. Es bastante claro que la manera en que se midan estas variables no tiene nada que ver con qué tan fuerte sea la relación entre ellas. Sin embargo, la covarianza entre ellas sí depende de las unidades de medición.

El hecho de que la covarianza dependa de las unidades de medición es una deficiencia que se supera mediante el **coeficiente de correlación** entre X y Y:

$$\text{Corr}(X,Y) \equiv \frac{\text{Cov}(X,Y)}{\text{de}(X)\cdot\text{de}(Y)} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y}; \qquad \boxed{\textbf{B.29}}$$

el coeficiente de correlación entre X y Y también suele denotarse ρ_{XY} (y se le suele llamar *correlación poblacional*).

Como σ_X y σ_Y son positivos, $\text{Cov}(X,Y)$ y $\text{Corr}(X,Y)$ siempre tienen el mismo signo y $\text{Corr}(X,Y) = 0$ si y solo si $\text{Cov}(X,Y) = 0$. Algunas de las propiedades de la covarianza se transfieren a la correlación. Si X y Y son independientes entonces $\text{Corr}(X,Y) = 0$, pero que la correlación sea cero no implica independencia. (Como la covarianza, también el coeficiente de correlación es una medida de dependencia lineal.) Sin embargo, la magnitud del coeficiente de correlación es más fácil de interpretar que la magnitud de la covarianza debido a las propiedades siguientes.

Propiedad CORR.1: $-1 \leq \text{Corr}(X,Y) \leq 1$.

Si $\text{Corr}(X,Y) = 0$, o lo que es equivalente $\text{Cov}(X,Y) = 0$, entonces no hay relación lineal entre X y Y, y se dice que X y Y son **variables aleatorias no correlacionadas**; de lo contrario, se dice que X y Y están *correlacionadas*. $\text{Corr}(X,Y) = 1$ implica una relación lineal positiva perfecta, lo que significa que se puede escribir $Y = a + bX$ para alguna constante a y alguna constante $b > 0$. $\text{Corr}(X,Y) = -1$ implica una relación lineal negativa perfecta, de manera que $Y = a + bX$ para alguna $b < 0$. Rara vez se presentan los casos extremos de 1 positivo o 1 negativo. Los valores de ρ_{XY} cercanos a 1 o a -1 indican una fuerte relación lineal.

Como se dijo antes, la correlación entre X y Y es invariante a las unidades de medición de X y de Y. Esto se enuncia de manera más general como sigue.

Propiedad CORR.2: Para todas las constantes a_1, b_1, a_2 y b_2, con $a_1a_2 > 0$,

$$\text{Corr}(a_1X + b_1, a_2Y + b_2) = \text{Corr}(X,Y).$$

Si $a_1a_2 < 0$, entonces

$$\text{Corr}(a_1X + b_1, a_2Y + b_2) = -\text{Corr}(X,Y).$$

Suponga, por ejemplo, que en la población trabajadora, la correlación entre ingreso y educación es .15. Esta medida no depende de si los ingresos se miden en dólares, miles de dólares o en cualquier otra unidad; tampoco depende de si le educación se mide en años, trimestres, meses, etcétera.

Varianza de sumas de variables aleatorias

Una vez que se han definido covarianza y correlación, se puede continuar con la lista de propiedades de la varianza.

Propiedad VAR.3: Para todas las constantes a y b,

$$\mathrm{Var}(aX + bY) = a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y) + 2ab\mathrm{Cov}(X,Y).$$

Se sigue inmediatamente que, si X y Y no están correlacionadas —de manera que $\mathrm{Cov}(X,Y) = 0$— entonces

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) \qquad \boxed{\mathbf{B.30}}$$

y

$$\mathrm{Var}(X - Y) = \mathrm{Var}(X) + \mathrm{Var}(Y). \qquad \boxed{\mathbf{B.31}}$$

En el último caso observe que la varianza de las diferencias es igual a la *suma de las varianzas*, no a su diferencia.

Como ejemplo de (B.30), sea X los beneficios obtenidos en un restaurant en una noche de viernes y sea Y los beneficios obtenidos en la noche del sábado siguiente. Entonces, $Z = X + Y$ son los beneficios en las dos noches. Suponga que X y Y tienen, cada una, un valor esperado de \$300 y una desviación estándar de \$15 (de manera que su varianza es 225). Los beneficios esperados por las dos noches es $\mathrm{E}(Z) = \mathrm{E}(X) + \mathrm{E}(Y) = 2\cdot(300) = 600$ dólares. Si X y Y son independientes, y lo tanto no están correlacionadas, entonces la varianza de los beneficios totales es la suma de las varianzas: $\mathrm{Var}(Z) = \mathrm{Var}(X) + \mathrm{Var}(Y) = 2\cdot(225) = 450$. Entonces, la desviación estándar de los beneficios totales es $\sqrt{450}$ es decir, aproximadamente \$21.21.

Las expresiones (B.30) y (B.31) pueden extenderse a más de dos variables aleatorias, para esta extensión se necesita una definición. Las variables aleatorias $\{X_1, \ldots, X_n\}$ son **variables aleatorias no correlacionadas entre sí**, si ninguna de las variables del conjunto está correlacionada con ninguna de las otras variables del conjunto. Es decir, $\mathrm{Cov}(X_i,X_j) = 0$, para toda $i \neq j$.

Propiedad VAR.4: Si $\{X_1, \ldots, X_n\}$ son variables aleatorias no correlacionadas entre sí y $\{a_i: i = 1, \ldots, n\}$ son constantes, entonces

$$\mathrm{Var}(a_1X_1 + \ldots + a_nX_n) = a_1^2\mathrm{Var}(X_1) + \ldots + a_n^2\mathrm{Var}(X_n).$$

Empleando la sumatoria, esto puede expresarse como

$$\mathrm{Var}\left(\sum_{i=1}^{n} a_iX_i\right) = \sum_{i=1}^{n} a_i^2\mathrm{Var}(X_i). \qquad \boxed{\mathbf{B.32}}$$

Un caso especial de la propiedad VAR.4 se presenta cuando se toma $a_i = 1$ para toda i. Entonces, para variables aleatorias no correlacionadas entre sí la varianza de la suma es igual a la suma de las varianzas:

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i). \qquad \boxed{\mathbf{B.33}}$$

Como las variables aleatorias independientes no están correlacionadas (vea la propiedad COV.1), la varianza de una suma de variables aleatorias dependientes es igual a la suma de las varianzas.

Si las X_i son variable no correlacionadas entre sí, entonces la expresión para $\mathrm{Var}\left(\sum_{i=1}^{n} a_iX_i\right)$ es mucho más complicada; en el lado derecho de (B.32) es necesario agregar los términos $2a_i\, a_j\mathrm{Cov}(x_i, x_j)$ para toda $i > j$.

Empleando la ecuación (B.33) se puede obtener la varianza de una variable aleatoria binomial. Sea $X \sim \text{Binomial}(n,\theta)$ si se escribe $X = Y_1 + \ldots + Y_n$, donde las Y_i son variables aleatorias independientes Bernoulli(θ). Entonces, de acuerdo con (B.33), $\text{Var}(X) = \text{Var}(Y_1) + \ldots + \text{Var}(Y_n) = n\theta(1 - \theta)$.

En el ejemplo de las reservaciones de la aerolínea, para $n = 120$ y $\theta = .85$, la varianza en la cantidad de pasajeros que utilicen su reservación es $120(.85)(.15) = 15.3$, de manera que la desviación estándar es 3.9.

Esperanza condicional

La covarianza y la correlación miden la relación lineal entre dos variables aleatorias tratándolas simétricamente. Pero en las ciencias sociales, es frecuente que se desee explicar una variable Y, en términos de otra, X. Además, si Y está relacionada con X en un forma que no sea lineal, también se desea saber esto. Sea Y la variable explicada y X la variable explicativa. Por ejemplo, Y puede ser el salario por hora y X los años de educación formal.

Ya antes se introdujo el concepto de una función de densidad de probabilidad condicional de Y dada X. Suponga que se quiere ver cómo cambia la distribución de los salarios con base en el nivel de educación. Sin embargo, por lo general se quiere encontrar una manera sencilla de resumir esta distribución. En este caso, un solo número no podrá ser suficiente, ya que la distribución de Y dado $X = x$ depende, por lo general, del valor de x. No obstante, la relación entre Y y X puede resumirse empleando la *esperanza condicional* de Y dado X, que también suele llamarse *media condicional*. La idea es esta: suponga que se sabe que X ha tomado un determinado valor, por ejemplo, x. Entonces, se puede calcular el valor esperado de Y, dado que se conoce este valor que ha tomado X. Este valor esperado se denota $\text{E}(Y|X = x)$, o también $\text{E}(Y|x)$ para abreviar. En general, cuando x cambia, también lo hace $\text{E}(Y|x)$.

Si Y es una variable aleatoria discreta que toma los valores $\{y_1, \ldots, y_m\}$, entonces

$$\text{E}(Y|x) = \sum_{j=1}^{m} y_j f_{Y|X}(y_j|x).$$

Si Y es continua, $\text{E}(Y|x)$ se define integrando $yf_{Y|X}(y|x)$ sobre todo los valores posibles de y. Como ocurre con las esperanzas no condicionales, la esperanza condicional es un promedio ponderado de los posibles valores de Y, pero ahora los pesos de ponderación reflejan el hecho de que X ha tomado un determinado valor. Por tanto, $\text{E}(Y|x)$ es una función de x, que dice cómo varía el valor esperado de Y de acuerdo con x.

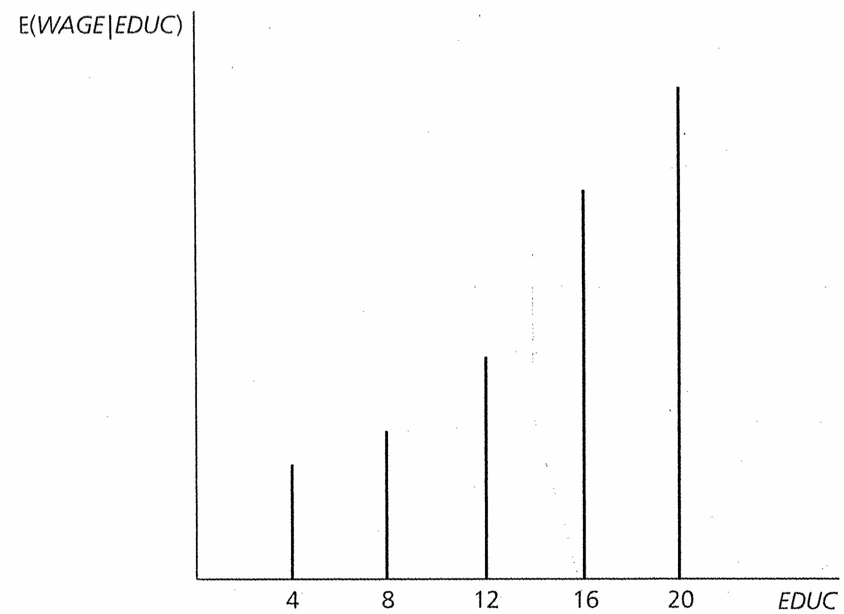
Por ejemplo, sea (X, Y) la población de todos los individuos que trabajan, donde X es años de educación y Y es salario por hora. Entonces, $\text{E}(Y|X = 12)$ es el salario promedio por hora de toda la población que tiene 12 años de educación (más o menos hasta bachillerato). $\text{E}(Y|X = 16)$ es el salario promedio por hora de todas las personas que cuentan con 16 años de educación. Los valores esperados correspondientes a varios niveles de educación proporcionan información importante sobre la relación entre salarios y educación. Véase la figura B.5.

En principio, pueden encontrarse los valores esperados de los salarios por hora correspondientes a cada nivel de educación y estas esperanzas pueden resumirse en una tabla. Pero dado que la educación puede variar mucho —e incluso puede medirse hasta en fracciones de años— esta es una forma muy complicada de medir la relación entre salario promedio y cantidad de educación. En econometría, lo que se acostumbra hacer es dar funciones sencillas que expresen esta relación. Por ejemplo, suponga que el valor esperado de *SALARIO* dada *EDUC* es una función lineal

$$\text{E}(WAGE|EDUC) = 1.05 + .45\,EDUC.$$

FIGURA B.5

Valor esperado del salario por hora dados diversos niveles de educación.



Si esta relación se satisface en la población de las personas que trabajan, el salario promedio de una persona que tenga 8 años de educación es 1.05 + .45(8) = 4.65, o \$4.65, es decir, \$4.65. El salario promedio de una persona que tenga 16 años de educación será 8.25, es decir, \$8.25. El coeficiente de la variable *EDUC* implica que por cada año más de educación el salario esperado por hora aumenta .45, es decir 45¢.

Las esperanzas condicionales también pueden ser funciones no lineales. Por ejemplo, suponga que $E(Y|x) = 10/x$, donde X es una variable aleatoria que es siempre mayor que cero. Esta función se grafica en la figura B.6 y puede representar la función de demanda, donde Y es cantidad demandada y X es precio. Si Y y X están relacionadas de esta manera, un análisis de asociación lineal, como el de correlación, será un análisis incompleto.

Propiedades de la esperanza condicional

En el análisis econométrico hay varias propiedades de las esperanzas condicionales que son muy útiles.

Propiedad CE.1: $E[c(X)|X] = c(X)$, para toda función $c(X)$.

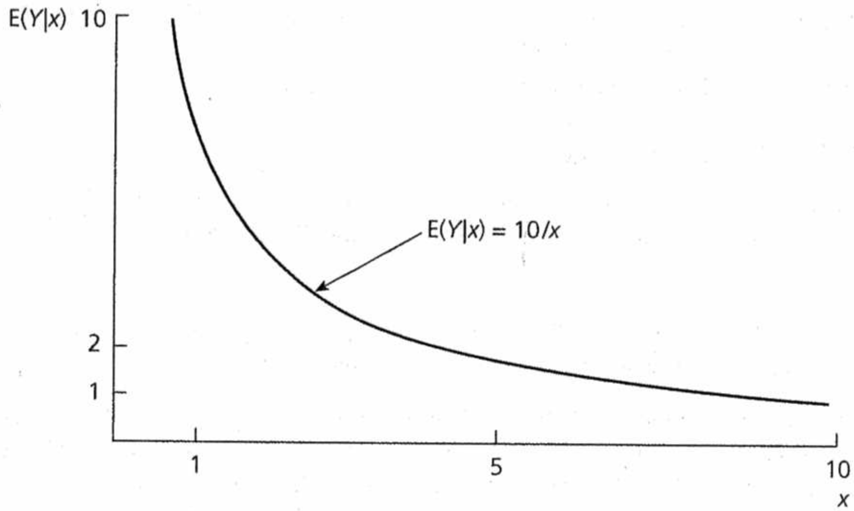
Esta primera propiedad significa que las funciones de X se comportan como constantes al calcular la esperanza condicional dado X. Por ejemplo, $E(X^2|X) = X^2$. Intuitivamente esto significa que si se conoce X, también se conoce X^2.

Propiedad CE.2: Dadas las funciones $a(X)$ y $b(X)$,

$$E[a(X)Y + b(X)|X] = a(X)E(Y|X) + b(X).$$

FIGURA B.6

Gráfica de $E(Y|x) = 10/x$.



Por ejemplo, la esperanza condicional de una función como $XY + 2X^2$: $E(XY + 2X^2|X) = XE(Y|X) + 2X^2$ puede calcularse fácilmente.

En la propiedad siguiente se relacionan los conceptos de independencia y esperanza condicional.

Propiedad CE.3: Si X y Y son independientes, entonces $E(Y|X) = E(Y)$.

Esta propiedad dice que, si X y Y son independientes, entonces el valor esperado de Y dado X no depende de X, en cuyo caso, $E(Y|X)$ es siempre igual al valor esperado (incondicional) de Y. En el ejemplo del salario y la educación, si el salario es independiente de la educación, entonces el salario promedio de las personas con bachillerato y con universidad será el mismo. Ya que esto es falso, no se puede suponer que salario y educación sean independientes.

Un caso especial de la propiedad EC.3 es el siguiente: Si U y X son independientes y $E(U) = 0$, entonces $E(U|X) = 0$.

Hay también propiedades de la esperanza condicional que tienen que ver con el hecho de que $E(Y|X)$ sea una función de X, por ejemplo, $E(Y|X) = \mu(X)$. Como X es una variable aleatoria, $\mu(X)$ es también una variable aleatoria. Además, $\mu(X)$ tiene una distribución de probabilidad y por tanto un valor esperado. En general, el valor esperado de $\mu(X)$ puede ser muy difícil de calcular de manera directa. La **ley de las esperanzas iteradas** dice que el valor esperado de $\mu(X)$ es simplemente igual al valor esperado de Y. Esto se expresa como sigue.

Propiedad CE.4: $E[E(Y|X)] = E(Y)$.

A primera vista esta propiedad parece un poco difícil de entender. Significa que, si se obtienen primero $E(Y|X)$ como función de X y de esto se determina el valor esperado (por supuesto, respecto la distribución de X, entonces se acaba obteniendo $\mathrm{E}(Y)$. Esto no es muy obvio, pero se puede deducir usando la definición de valor esperado.

Un ejemplo de cómo usar la propiedad EC.4, sea $Y = SALARIO$ y $X = EDUC$, donde *SALARIO* se mide en horas y *EDUC*, en años. Suponga que el valor esperado de *SALARIO* dado *EDUC* es $\mathrm{E}(SALARIO|EDUC) = 4 + .60\ EDUC$. Además, $\mathrm{E}(EDUC) = 11.5$. Entonces, la ley de las esperanzas iteradas implica que $\mathrm{E}(SALARIO) = \mathrm{E}(4 + .60\ EDUC) = 4 + .60\ \mathrm{E}(EDUC) = 4 + .60(11.5) = 10.90$, es decir, \$10.90 por hora.

La propiedad siguiente da una versión más general de la ley de las esperanzas iteradas.

Propiedad CE.4′: $\mathrm{E}(Y|X) = \mathrm{E}[\mathrm{E}(Y|X,Z)|X]$.

En otras palabras, $\mathrm{E}(Y|X)$ se puede obtener en dos pasos. Primero, se encuentra $\mathrm{E}(Y|X,Z)$ para cualquier otra variable aleatoria Z. Después se encuentra el valor esperado de $\mathrm{E}(Y|X,Z)$, condicionado a X.

Propiedad CE.5: Si $\mathrm{E}(Y|X) = \mathrm{E}(Y)$, entonces $\mathrm{Cov}(X,Y) = 0$ [y entonces $\mathrm{Corr}(X,Y) = 0$]. De hecho, *toda* función de X es una función no correlacionada con Y.

Esta propiedad significa que, si el conocer X no modifica el valor esperado de Y, entonces X y Y no *deben* estar correlacionadas, lo que implica que si X y Y están correlacionadas, entonces $\mathrm{E}(Y|X)$ debe depender de X. El inverso de la propiedad EC.5 no es verdad: si X y Y no están correlacionadas, entonces $\mathrm{E}(Y|X)$ *puede* depender de X. Por ejemplo, suponga que $Y = X^2$. Entonces, $\mathrm{E}(Y|X) = X^2$, lo que es claramente una función de X. Sin embargo, como se dijo al hablar de la covarianza y de la correlación, es posible que X y X^2 no estén correlacionadas. La esperanza condicional capta las relaciones no lineales entre X y Y que el análisis de correlación no puede captar.

Las propiedades EC.4 y EC.5 tienen dos consecuencias importantes: si U y X son variables aleatorias tales que $\mathrm{E}(U|X) = 0$, entonces $\mathrm{E}(U) = 0$ y U y X *no* están correlacionadas.

Propiedad CE.6: Si $\mathrm{E}(Y^2) < \infty$ y $\mathrm{E}[g(X)^2] < \infty$ para alguna función g, entonces $\mathrm{E}\{[Y - \mu(X)]^2|X\} \leq \mathrm{E}\{[Y - g(X)]^2|X\}$ y $\mathrm{E}\{[Y - \mu(X)]^2\} \leq \mathrm{E}\{[Y - g(X)]^2\}$.

La propiedad EC.6 es muy útil en contextos de predicción o de pronóstico. La primera desigualdad indica que si se mide la inexactitud de una predicción como el cuadrado del error de predicción *esperado*, condicionado a X, entonces, para predecir X la media condicional es mejor que cualquier otra función de Y. La media condicional también minimiza el cuadrado del error de predicción esperado no condicional.

Varianza condicional

Dadas las variables aleatorias X y Y, la varianza de Y, condicionada a $X = x$, es simplemente la varianza asociada a la distribución condicional de Y, dado $X = x$: $\mathrm{E}\{[Y - \mathrm{E}(Y|x)]^2|x\}$. La fórmula

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

suele usarse para hacer cálculos. Sólo ocasionalmente habrá que calcular una varianza condicional. Pero para ciertos temas del análisis de regresión se tendrá que usar la varianza condicional y hacer supuestos acerca de ella.

Por ejemplo, sea $Y = AHORROS$ y $X = INGRESO$ (medidos ambos anualmente en la población de todas las familias). Suponga que $\mathrm{Var}(AHORROS|INGRESO) = 400 + .25\ INGRESO$. Esto indica que a medida que aumenta el ingreso, aumenta la varianza en los niveles de ahorro.

Es importante notar que la relación entre la varianza del *AHORROS* y el *INGRESO* es totalmente distinta a la relación entre *valor esperado* de los *AHORROS* e *INGRESO*.

A continuación se enuncia una útil propiedad acerca de la varianza condicional.

Propiedad CV.1: Si X y Y son independientes, entonces $\mathrm{Var}(Y|X) = \mathrm{Var}(Y)$.

Esta propiedad es bastante clara, ya que la distribución de Y dado X no depende de X y $\mathrm{Var}(Y|X)$ es sólo una característica de esta distribución.

B.5 La distribución normal y otras distribuciones semejantes

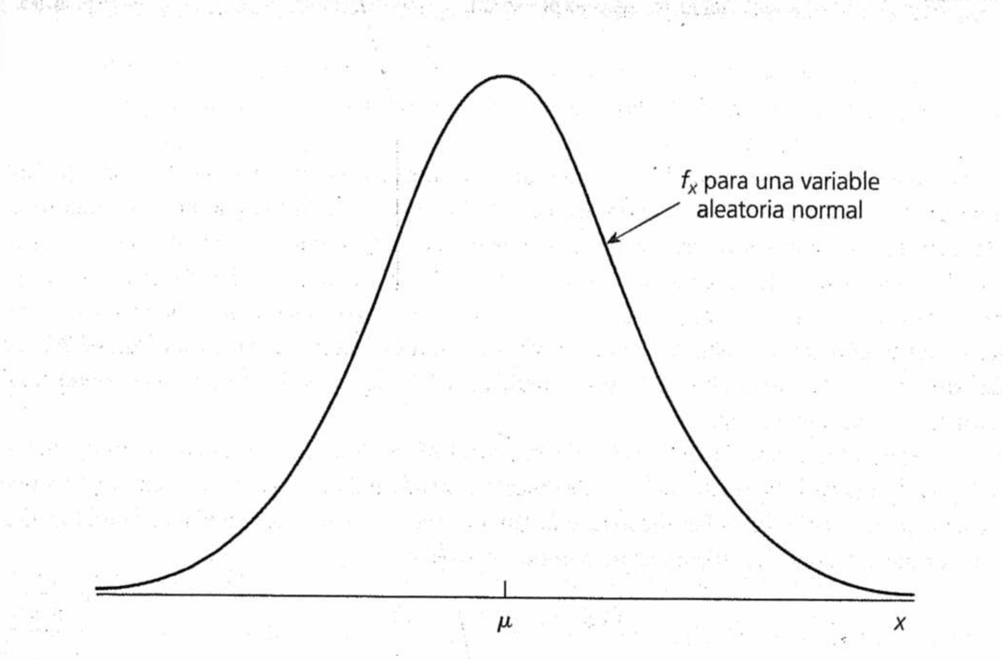
La distribución normal

La distribución normal y las distribuciones que de ella se derivan son las más empleadas en estadística y en econometría. El supuesto de que una variable aleatoria, definida para toda una población, tiene una distribución normal simplifica el cálculo de probabilidades. Además, la distribución normal y las distribuciones semejantes son muy usadas en la estadística y en la econometría para hacer inferencias —aun cuando la población subyacente no sea necesariamente normal. Por ahora, se deben posponer los detalles, pero con toda seguridad este tipo de distribuciones aparecerán con frecuencia en todo este libro.

Una variable aleatoria normal es una variable aleatoria continua que puede tomar cualquier valor. Su función de densidad de probabilidad tiene la conocida forma de curva de campana que se muestra en la figura B.7.

FIGURA B.7

Forma general de la función de densidad de probabilidad normal.



En términos matemáticos, la fdp de X se expresa como

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp[-(x - \mu)^2/2\sigma^2], \quad -\infty < x < \infty, \qquad \textbf{B.34}$$

donde $\mu = \mathrm{E}(X)$ y $\sigma^2 = \mathrm{Var}(X)$. Se dice que X tiene una **distribución normal** cuyo valor esperado es μ y cuya varianza es σ^2, y se escribe $X \sim \mathrm{Normal}(\mu, \sigma^2)$. Como la distribución normal es simétrica respecto a μ, μ es también la mediana de X. A la distribución normal también se le suele llamar *distribución gaussiana* en honor al famoso matemático C. F. Gauss.

Algunas variables aleatorias parecen seguir poco más o menos una distribución normal. Las estaturas y los pesos de los seres humanos, las puntuaciones de test y las tasas de desempleo en los condados (de Estados Unidos) tienen una fdp cuya forma es más o menos la de la figura B.7. Hay otras distribuciones, como la del ingreso, que no siguen una función de probabilidad normal. En la mayoría de los países, el ingreso no está distribuido simétricamente respecto a ningún valor; esta distribución es sesgada hacia la cola superior. En algunos casos, una variable puede transformarse para tener normalidad. Una transformación usual es el logaritmo natural, que sólo tiene sentido para variables aleatorias positivas. Si X es una variable aleatoria positiva, como ingreso, y $Y = \log(X)$ tiene una distribución normal, entonces se dice que X tiene una *distribución lognormal*. Se ha encontrado que en muchos países la distribución lognormal se ajusta bastante bien a la distribución del ingreso. Hay otras variables, como los precios de los bienes, que también pueden describirse como distribuidas lognormales.

La distribución normal estándar

Un caso especial de distribución normal es la distribución normal cuya media es cero y cuya varianza (y, por tanto, desviación estándar) es uno. Si una variable aleatoria Z tiene una distribución normal (0,1) entonces se dice que sigue la **distribución normal estándar**. La fdp de una variable normal estándar se denota $\phi(z)$; de acuerdo con (B.34), dado que $\mu = 0$ y $\sigma^2 = 1$, esta función está dada por

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \exp(-z^2/2), \quad -\infty < z < \infty. \qquad \textbf{B.35}$$

La función de la distribución acumulada normal estándar se denota $\Phi(z)$ y es el área bajo ϕ, a la izquierda de z; vea la figura B.8. Hay que recordar que $\Phi(z) = \mathrm{P}(Z \leq z)$; dado que Z es continua, $\Phi(z) = \mathrm{P}(Z < z)$ también es continua.

Para obtener los valores de $\Phi(z)$ no hay ninguna fórmula sencilla que pueda usarse [debido a que $\Phi(z)$ es la integral de la función dada en (B.35), y a que esta integral no tiene una forma cerrada]. No obstante, los valores de $\Phi(z)$ pueden tabularse fácilmente; en la tabla G.1 del apéndice G se dan estos valores para valores de z entre -3.1 y 3.1 Si $z \leq -3.1$, $\Phi(z)$ es menor que .001, y si $z \geq 3.1$, $\Phi(z)$ es mayor que .999. La mayoría del software para estadística y econometría tienen comandos sencillos para calcular los valores de la fda normal estándar, de manera que para obtener las probabilidades correspondientes á cualquier valor de z se puede prescindir totalmente de las tablas impresas

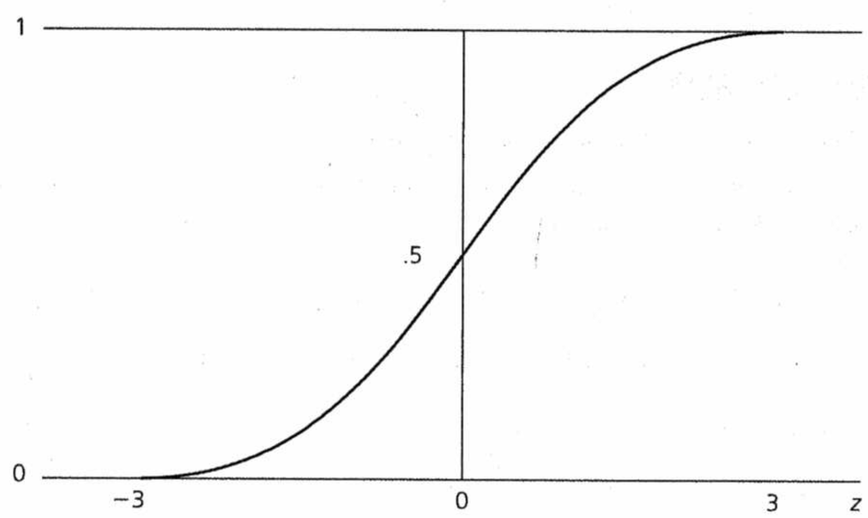
Empleando las propiedades básicas de la probabilidad —y, en particular, las propiedades (B.7) y (B.8) que se refieren a funciones de distribución acumulada— la fda normal estándar puede usarse para calcular la probabilidad de cualquier evento en el que intervenga una variable aleatoria normal estándar. Las fórmulas más importantes son

$$\mathrm{P}(Z > z) = 1 - \Phi(z), \qquad \textbf{B.36}$$

$$\mathrm{P}(Z < -z) = \mathrm{P}(Z > z), \qquad \textbf{B.37}$$

FIGURA B.8

Función de la distribución acumulada normal estándar.



y

$$P(a \leq Z \leq b) = \Phi(b) - \Phi(a). \qquad \textbf{B.38}$$

Como Z es una variable aleatoria continua, estas tres fórmulas son válidas, ya sea que las desigualdades sean o no estrictas. Algunos ejemplos son $P(Z > .44) = 1 - .67 = .33$, $P(Z < -.92) = P(Z > .92) = 1 - .821 = .179$ y $P(-1 < Z \leq .5) = .692 - .159 = .533$.

Otra expresión útil es que, para toda $c > 0$,

$$\begin{aligned} P(|Z| > c) &= P(Z > c) + P(Z < -c) \\ &= 2 \cdot P(Z > c) = 2[1 - \Phi(c)]. \end{aligned} \qquad \textbf{B.39}$$

Por tanto, la probabilidad de que el valor absoluto de Z sea mayor que alguna constante positiva c es simplemente el doble de la probabilidad $P(Z > c)$; esto refleja la simetría de la distribución normal estándar.

En la mayoría de las aplicaciones, se parte de una variable aleatoria distribuida normalmente, $X \sim \text{Normal}(\mu, \sigma^2)$, donde μ es diferente de cero y $\sigma^2 \neq 1$. Todas las variables aleatorias normales pueden transformarse en variables aleatorias normales estándar usando la siguiente propiedad.

Propiedad Normal.1: Si $X \sim \text{Normal}(\mu, \sigma^2)$, entonces $(X - \mu)/\sigma \sim \text{Normal}(0, 1)$.

La propiedad Normal.1 indica cómo transformar cualquier variable aleatoria normal en una variable aleatoria normal estándar. Así, suponga que $X \sim \text{Normal}(3, 4)$ y que se quiere calcular

$P(X \le 1)$. Entre los pasos para hacer esto está la normalización de la variable X a una variable normal estándar:

$$P(X \le 1) = P(X - 3 \le 1 - 3) = P\left(\frac{X-3}{2} \le -1\right)$$

$$= P(Z \le -1) = \Phi(-1) = .159.$$

Ejemplo B.6

[Probabilidades para una variable aleatoria normal]

Primero, se calculará $P(2 < X \le 6)$ dado que $X \sim \text{Normal}(4, 9)$ (donde usar $<$ o $\le$ es irrelevante debido a que X es una variable aleatoria continua). Ahora,

$$P(2 < X \le 6) = P\left(\frac{2-4}{3} < \frac{X-4}{3} \le \frac{6-4}{3}\right) = P(-2/3 < Z \le 2/3)$$

$$= \Phi(.67) - \Phi(-.67) = .749 - .251 = .498.$$

Ahora, se calculará $P(|X| > 2)$:

$$\begin{aligned} P(|X| > 2) &= P(X > 2) + P(X < -2) \\ &= P[(X-4)/3 > (2-4)/3] + P[(X-4)/3 < (-2-4)/3] \\ &= 1 - \Phi(-2/3) + \Phi(-2) \\ &= 1 - .251 + .023 = .772. \end{aligned}$$

Propiedades adicionales de la distribución normal

Esta subsección concluye dando algunas otras propiedades de las distribuciones normales, que se usarán más adelante.

Propiedad Normal.2: Si $X \sim \text{Normal}(\mu, \sigma^2)$, entonces $aX + b \sim \text{Normal}(a\mu + b, a^2\sigma^2)$.

Por tanto, si $X \sim \text{Normal}(1, 9)$, entonces $Y = 2X + 3$ tiene una distribución normal con media $2E(X) + 3 = 5$ y varianza $2^2 \cdot 9 = 36$; $sd(Y) = 2sd(X) = 2 \cdot 3 = 6$.

Ya antes se discutió cómo, en general, correlación cero e independencia no son lo mismo. En el caso de variables aleatorias distribuidas normalmente, resulta que correlación cero es suficiente para independencia.

Propiedad Normal.3: Si X y Y son variables conjuntas distribuidas normalmente, entonces X y Y son independientes si y solo si, $\text{Cov}(X, Y) = 0$.

Propiedad Normal.4: Toda combinación lineal de variables aleatorias normales idénticamente distribuidas e independientes tiene una distribución normal.

Por ejemplo, sean X_i, para $i = 1, 2$ y 3, variables aleatorias independientes con distribución $\text{Normal}(\mu, \sigma^2)$. Se define $W = X_1 + 2X_2 - 3X_3$. Entonces, W está distribuida normalmente; sólo hace falta hallar su media y su varianza. Ahora,

$$E(W) = E(X_1) + 2E(X_2) - 3E(X_3) = \mu + 2\mu - 3\mu = 0.$$

y,

$$\text{Var}(W) = \text{Var}(X_1) + 4\text{Var}(X_2) + 9\text{Var}(X_3) = 14\sigma^2.$$

La propiedad Normal.4 también implica que el promedio de variables aleatorias normalmente distribuidas e independientes tiene una distribución normal. Si $Y_1, Y_2, \ldots, Y_n$ son variables aleatorias independientes y cada una tiene la distribución Normal(μ, σ^2), entonces

$$\bar{Y} \sim \text{Normal}(\mu, \sigma^2/n). \qquad \textbf{B.40}$$

Este resultado es crucial en la inferencia estadística acerca de la media de una población normal.

Hay otras propiedades de la distribución normal que vale la pena conocer, aunque no tengan un papel importante en este libro. Como una variable aleatoria normal es simétrica respecto a su media, tiene sesgo cero, es decir, $\text{E}[(X - \mu)^3] = 0$. Además, se puede demostrar que

$$\text{E}[(X - \mu)^4]/\sigma^4 = 3,$$

o $\text{E}(Z^4) = 3$, donde Z tiene distribución normal estándar. Como la distribución normal se encuentra con tanta frecuencia en probabilidad y estadística, la medida de la curtosis de cualquier variable aleatoria X (cuyo cuarto momento exista) suele definirse como $\text{E}[(X - \mu)^4]/\sigma^4 - 3$, es decir, relativo al valor de la distribución normal estándar. Si $\text{E}[(X - \mu)^4]/\sigma^4 > 3$, entonces la distribución de X tienen colas más gruesas que la distribución normal (algo que suele ocurrir, como por ejemplo en la distribución t, que se presentará en breve); si $\text{E}[(X - \mu)^4]/\sigma^4 < 3$, entonces la distribución tienen colas más delgadas que la normal (una situación poco frecuente).

La distribución *ji*-cuadrada

La distribución *ji*-cuadrada se obtiene directamente a partir de variables normales estándar, independientes. Sean Z_i, $i = 1, 2, \ldots, n$, variables aleatorias independientes, cada una con una distribución normal estándar. Se define una nueva variable como la suma de los cuadrados de las Z_i:

$$X = \sum_{i=1}^{n} Z_i^2. \qquad \textbf{B.41}$$

Entonces, X tiene lo que se conoce como una **distribución *ji*-cuadrada** con n **grados de libertad** (*gl*). Esto se expresa $X \sim \chi_n^2$. En una distribución *ji*-cuadrada, los *gl* corresponden a la cantidad de términos en la suma de la ecuación (B.14). En este libro, el concepto de grados de libertad será muy importante en los análisis estadísticos y econométricos.

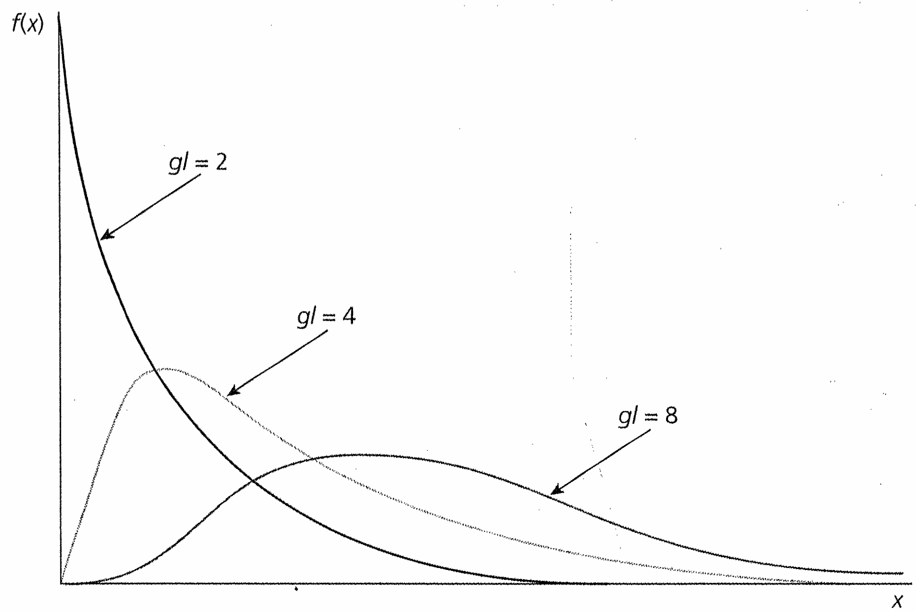
En la figura B.9 se presentan las fdp de distribuciones *ji*-cuadrada correspondientes a diversos grados de libertad; la fórmula de esta fdp no se necesitará, por lo que no se reproduce aquí. De acuerdo con la ecuación (B.41), es claro que las variables aleatorias *ji*-cuadrada son siempre no negativas, y que, a diferencia de la distribución normal, la distribución *ji*-cuadrada no es simétrica respecto a ningún punto. Se puede demostrar que es si $X \sim \chi_n^2$, entonces el valor esperado de X es n [el número de términos en (B.41)] y la varianza de X es $2n$.

La distribución *t*

La distribución t es el caballo de batalla de la estadística clásica y del análisis de regresión múltiple. Una distribución t se obtiene a partir de una variable aleatoria normal estándar y una variable aleatoria *ji*-cuadrada.

FIGURA B.9

Distribuciones *ji*-cuadrada para diversos grados de libertad.



Sea Z una variable aleatoria que tiene una distribución normal estándar y sea X una variable aleatoria que tiene una distribución *ji*-cuadrada con n grados de libertad. Suponga, además, que Z y X son independientes. Entonces, la variable aleatoria

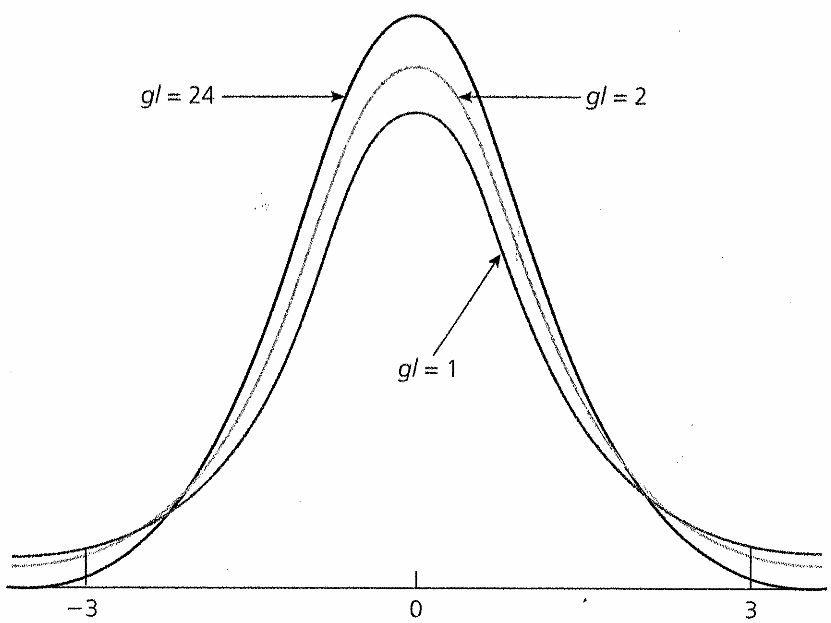
$$T = \frac{Z}{\sqrt{X/n}} \tag{B.42}$$

tiene una **distribución *t*** con n grados de libertad. Esto se denotará $T \sim t_n$. Las distribuciones t obtienen sus grados de libertad de la variable aleatoria *ji*-cuadrada en el denominador de la ecuación (B.42).

La fdp de la distribución t tienen una forma similar a la de la distribución normal estándar, sólo que es más dispersa, y por tanto tiene áreas mayores en las colas. El valor esperado de una variable aleatoria con distribución t es cero (estrictamente hablando, el valor esperado existe sólo para $n > 1$) y la varianza es is $n/(n-2)$ para $n > 2$. (Para $n \leq 2$ no existe la varianza debido a que la distribución es demasiado dispersa.) En la figura B.10 se grafican distribuciones t correspondientes a diversos grados de libertad. A medida que los grados de libertad aumentan, las distribuciones t se aproximan a la distribución normal estándar.

FIGURA B.10

Distribuciones t con diversos grados de libertad.



La distribución F

Otra distribución importante para aquellos que se dedican a la estadística o a la econometría es la distribución F. Ésta se usará en especial para probar hipótesis en el contexto del análisis de regresión múltiple.

Para definir una variable aleatoria F, sean $X_1 \sim \chi^2_{k_1}$ y $X_2 \sim \chi^2_{k_2}$ y suponga que X_1 y X_2 son independientes. Entonces, la variable aleatoria

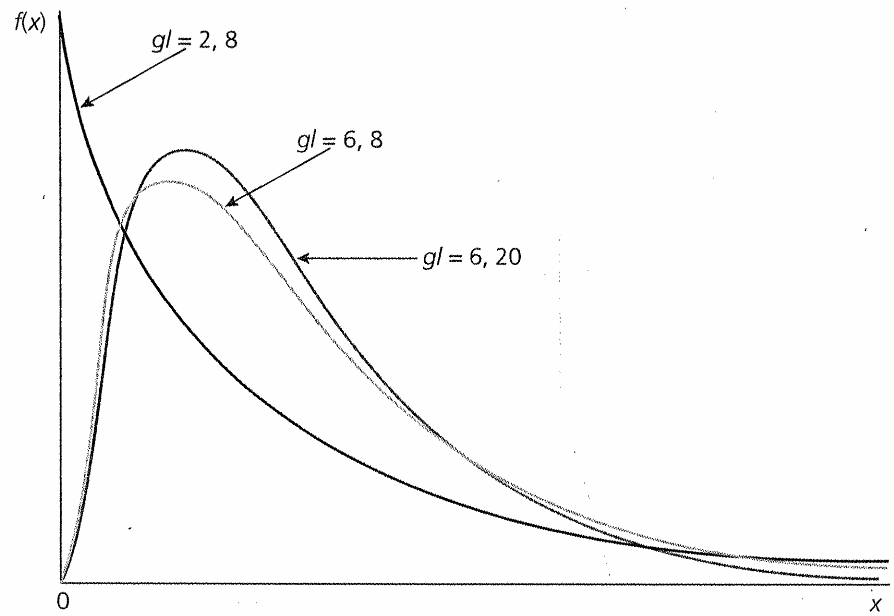
$$F = \frac{(X_1/k_1)}{(X_2/k_2)} \quad \textbf{B.43}$$

tiene una **distribución F** con (k_1, k_2) grados de libertad. Esto se denota como $F \sim F_{k_1, k_2}$. En la figura B.11 se muestran las fdp de distribuciones F con diversos grados de libertad.

En F_{k_1, k_2} el orden de los grados de libertad es crítico. El entero k_1 son los *grados de libertad en el numerador*, debido a que corresponde a la variable *ji*-cuadrada del numerador. De igual manera, el entero k_2 son los *grados de libertad en el denominador*, debido a que corresponde a la variable *ji*-cuadrada del denominador. Aquí hay que tener cuidado, pues la ecuación (B.34) también puede expresarse como $(X_1k_2)/(X_2k_1)$, de manera que k_1 aparece en el denominador. Sólo hay que recordar que los *gl* del numerador es el entero asociado con la variable *ji*-cuadrada en el numerador de la ecuación (B.43) y de manera similar para *gl del denominador*.

FIGURA B.11

Distribuciones F_{k_1, k_2} para diversos grados de libertad, k_1 y k_2.



RESUMEN

En este apéndice se hizo un repaso de los conceptos de probabilidad que se emplean en la econometría. La mayoría de ellos resultará familiar al lector por los cursos introductorios de probabilidad y estadística. Algunos de los temas más avanzados, por ejemplo las propiedades de las esperanzas condicionales, no es necesario que se dominen ahora —habrá tiempo para eso cuando estos conceptos surjan en el contexto del análisis de regresión, en la parte 1.

En un curso introductorio de estadística, la atención se centra en los cálculos de medias, varianzas, covarianzas, etcétera de las diferentes distribuciones. En la parte 1 no se necesitarán estos cálculos: lo que se usará serán las *propiedades* de la esperanza, varianza, etcétera, vistas en este apéndice.

TÉRMINOS CLAVE

Coeficiente de correlación
Covarianza
Curtosis
Desviación estándar
Distribución binomial
Distribución *ji*-cuadrada
Distribución condicional
Distribución conjunta
Distribución *F*
Distribución normal
Distribución normal estándar
Distribución simétrica
Distribución *t*
Esperanza condicional
Experimento

Función de densidad de probabilidad (fdp)
Función de distribución acumulada (fda)
Grados de libertad
Ley de las esperanzas iteradas
Mediana
Sesgo
Valor esperado
Variable aleatoria
Variable aleatoria continua
Variable aleatoria de Bernoulli (o binaria)
Variable aleatoria discreta
Variable aleatoria estandarizada
Variables aleatorias independientes
Variables aleatorias no correlacionadas
Variables aleatorias no correlacionadas entre sí
Varianza

PROBLEMAS

B.1 Suponga que un estudiante de bachillerato se está preparando para presentar el examen SAT de admisión a la universidad. Explique por qué la calificación que obtendrá en este examen puede ser considerada como una variable aleatoria.

B.2 Sea X una variable aleatoria con distribuciones normal (5, 4). Determine las probabilidades de los eventos siguientes:

i) $P(X \leq 6)$.

ii) $P(X > 4)$.

iii) $P(|X - 5| > 1)$.

B.3 Se ha hablado mucho acerca de que hay algunos fondos mutualistas que superan el mercado año tras año (es decir, el rendimiento de las acciones de los fondos mutualistas es más alto que el de un portafolio como el S&P 500). En concreto, considere un periodo de 10 años y la población formada por los 4,170 fondos mutualistas publicados en *The Wall Street Journal* el 1 de enero de 1995. Decir que el desempeño en relación al mercado es aleatorio significa que cada año, cada fondo tiene 50 por ciento de posibilidades de superar el mercado y que su desempeño es independiente de un año a otro.

i) Si el desempeño en relación al mercado es realmente aleatorio, ¿cuál es la probabilidad de que un determinado fondo supere el mercado todos estos 10 años?

ii) Calcule la probabilidad de que en estos 10 años *por lo menos* uno de los 4,170 fondos supere el mercado?

iii) Empleando un software estadístico que calcule probabilidades binomiales, determine la probabilidad de que, en estos 10 años, por lo menos cinco fondos superen el mercado.

B.4 Dado un condado de Estados Unidos, tomado al azar, sea X la proporción de adultos mayores de 65 años que tienen un empleo, es decir, la tasa de empleo entre los adultos mayores. Entonces, X está restringida a tomar un valor entre cero y uno. Suponga que la función de distribución acumulada de X está dada por $F(x) = 3x^2 - 2x^3$ para $0 \leq x \leq 1$. Calcule la probabilidad de que la tasa de empleo entre los adultos mayores sea por lo menos .6 (60%).

B.5 Antes de elegir el jurado para el juicio por homicidio contra O. J. Simpson en 1995, en una encuesta se encontró que aproximadamente 20% de la población adulta creía que Simpson era inocente (dado que gran parte de las evidencias físicas del caso habían sido reveladas al público). Ignorando el hecho de que este 20% sea una estimación basada en una submuestra tomada de la población; considérese como ilustración, como el verdadero porcentaje de personas que, antes de la elección del jurado, pensaban que Simpson era inocente. Suponga que los 12 miembros del jurado fueron elegidos de la población de forma aleatoria e independiente (aunque esto resultó no ser cierto).

(i) Calcule la probabilidad de que entre los miembros del jurado haya habido por lo menos uno que antes de la elección del mismo haya creído en la inocencia de Simpson. [*Sugerencia:* defínase una variable aleatoria binomial (12, .20) X correspondiente a la cantidad de miembros del jurado que creían en la inocencia de Simpson.]

ii) Calcule la probabilidad de que entre los miembros del jurado haya habido al menos dos que hayan creído la inocencia de Simpson. [*Sugerencia:* $P(X \geq 2) = 1 - P(X \leq 1)$, y $P(X \leq 1) = P(X = 0) + P(X = 1)$.]

B.6 (Se requiere cálculo) Sea X los años de cárcel, en un determinado estado de Estados Unidos, para una persona acusada del robo de un automóvil. Suponga que la fdp de X sea

$$f(x) = (1/9)x^2, 0 < x < 3.$$

Use la integración para calcular los años de cárcel esperados.

B.7 Si un jugador de baloncesto tiene una probabilidad de 74% de anotar un tiro libre, entonces, en promedio, ¿cuántos tiros libres anotará en un juego en el que tiene ocho intentos de tiro libre?

B.8 Suponga que un estudiante universitario toma tres cursos: uno de dos créditos, otro de tres y otro más de cuatro. La calificación esperada en el curso de dos créditos es 3.5, mientras que la calificación esperada en los cursos de tres y cuatro créditos es 3.0. ¿Cuál es la calificación promedio esperada en el semestre? (Recuerde que cada calificación está ponderada por el número total de créditos.)

B.9 Sea X el salario anual de los profesores universitarios en Estados Unidos, medido en miles de dólares. Suponga que el sueldo promedio sea 52.3 y la desviación estándar 14.6. Halle la media y la desviación estándar si el sueldo se mide en dólares.

B.10 B.10 Suponga que en una universidad grande, el promedio de calificaciones obtenido, *GPA*, y la puntuación en el examen de admisión, *SAT*, se relacionan mediante la esperanza condicional $E(GPA|SAT) = .70 + .002\ SAT$.

i) Determine el *GPA* esperado cuando $SAT = 800$. Calcule $E(GPA|SAT = 1{,}400)$. Analice la diferencia.

ii) Si en esta universidad el *SAT* promedio es 1,100, ¿cuál será el *GPA* promedio? (*Sugerencia:* emplee la propiedad EC.4.)

iii) Si la puntuación obtenida en el *SAT* por un estudiante es 1,100, ¿significa esto que su calificación promedio, *GPA*, será la encontrada en el inciso ii)? Explique.