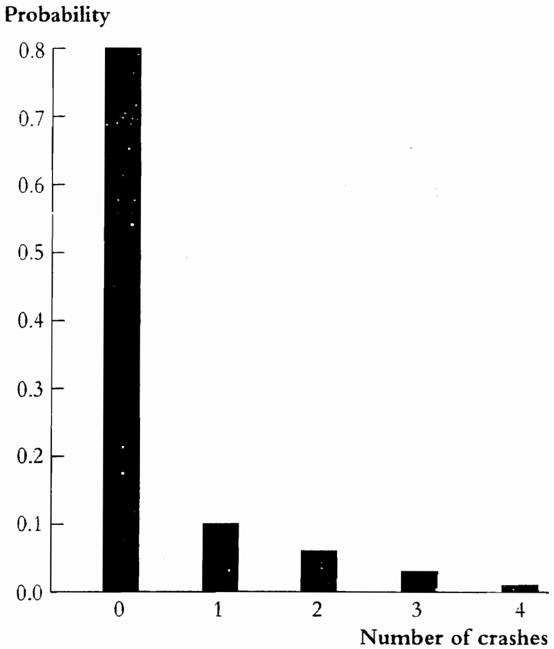
FIGURE 2.1 Probability Distribution of the Number of Computer Crashes

The height of each bar is the probability that the computer crashes the indicated number of times. The height of the first bar is 0.8, so the probability of 0 computer crashes is 80%. The height of the second bar is 0.1, so the probability of 1 computer crash is 10%, and so forth for the other bars.



A cumulative probability distribution is also referred to as a **cumulative distribution function**, a **c.d.f.**, or a **cumulative distribution.**

The Bernoulli distribution. An important special case of a discrete random variable is when the random variable is binary, that is, the outcomes are 0 or 1. A binary random variable is called a **Bernoulli random variable** (in honor of the seventeenth-century Swiss mathematician and scientist Jacob Bernoulli), and its probability distribution is called the **Bernoulli distribution.**

For example, let G be the gender of the next new person you meet, where $G = 0$ indicates that the person is male and $G = 1$ indicates that she is female. The outcomes of G and their probabilities thus are

$$G = \begin{cases} 1 \text{ with probability } p \\ 0 \text{ with probability } 1 - p, \end{cases} \tag{2.1}$$

where p is the probability of the next new person you meet being a woman. The probability distribution in Equation (2.1) is the Bernoulli distribution.

Probability Distribution of a Discrete Random Variable

Probability distribution. The **probability distribution** of a discrete random variable is the list of all possible values of the variable and the probability that each value will occur. These probabilities sum to 1.

For example, let M be the number of times your computer crashes while you are writing a term paper. The probability distribution of the random variable M is the list of probabilities of each possible outcome: The probability that $M = 0$, denoted $\Pr(M = 0)$, is the probability of no computer crashes; $\Pr(M = 1)$ is the probability of a single computer crash; and so forth. An example of a probability distribution for M is given in the second row of Table 2.1; in this distribution, if your computer crashes four times, you will quit and write the paper by hand. According to this distribution, the probability of no crashes is 80%; the probability of one crash is 10%; and the probability of two, three, or four crashes is, respectively, 6%, 3%, and 1%. These probabilities sum to 100%. This probability distribution is plotted in Figure 2.1.

Probabilities of events. The probability of an event can be computed from the probability distribution. For example, the probability of the event of one or two crashes is the sum of the probabilities of the constituent outcomes. That is, $\Pr(M = 1 \text{ or } M = 2) = \Pr(M = 1) + \Pr(M = 2) = 0.10 + 0.06 = 0.16$, or 16%.

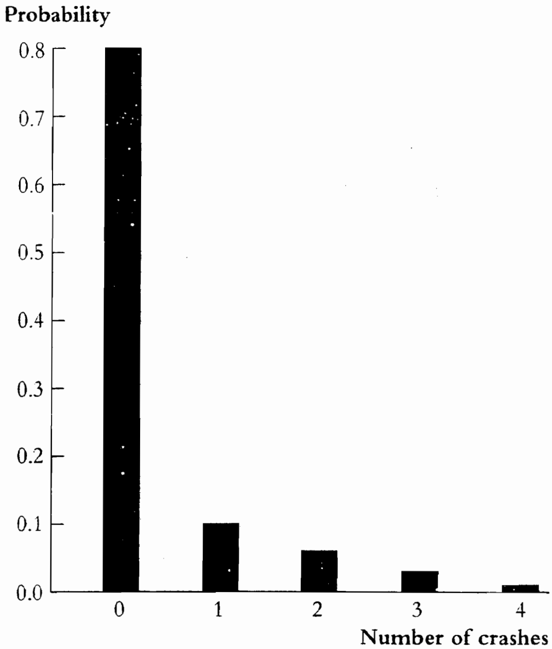
Cumulative probability distribution. The **cumulative probability distribution** is the probability that the random variable is less than or equal to a particular value. The last row of Table 2.1 gives the cumulative probability distribution of the random variable M. For example, the probability of at most one crash, $\Pr(M \leq 1)$, is 90%, which is the sum of the probabilities of no crashes (80%) and of one crash (10%).

TABLE 2.1 Probability of Your Computer Crashing *M* Times

	Outcome (number of crashes)				
	0	**1**	**2**	**3**	**4**
Probability distribution	0.80	0.10	0.06	0.03	0.01
Cumulative probability distribution	0.80	0.90	0.96	0.99	1.00

FIGURE 2.1 Probability Distribution of the Number of Computer Crashes

The height of each bar is the probability that the computer crashes the indicated number of times. The height of the first bar is 0.8, so the probability of 0 computer crashes is 80%. The height of the second bar is 0.1, so the probability of 1 computer crash is 10%, and so forth for the other bars.



A cumulative probability distribution is also referred to as a **cumulative distribution function**, a **c.d.f.**, or a **cumulative distribution.**

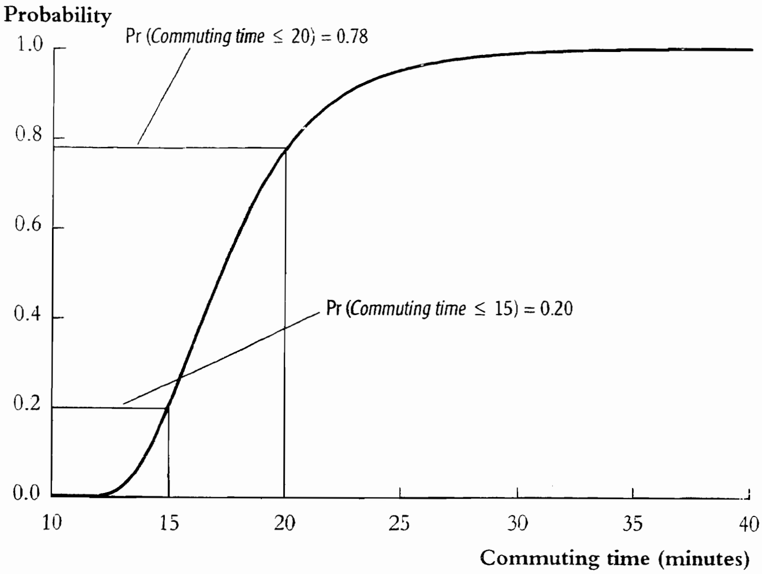
The Bernoulli distribution. An important special case of a discrete random variable is when the random variable is binary, that is, the outcomes are 0 or 1. A binary random variable is called a **Bernoulli random variable** (in honor of the seventeenth-century Swiss mathematician and scientist Jacob Bernoulli), and its probability distribution is called the **Bernoulli distribution.**

For example, let G be the gender of the next new person you meet, where $G = 0$ indicates that the person is male and $G = 1$ indicates that she is female. The outcomes of G and their probabilities thus are

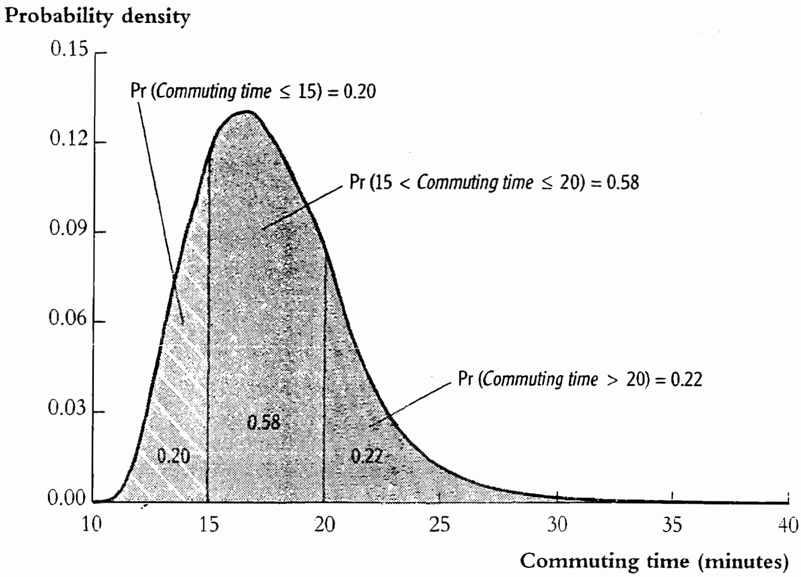
$$G = \begin{cases} 1 \text{ with probability } p \\ 0 \text{ with probability } 1 - p, \end{cases} \tag{2.1}$$

where p is the probability of the next new person you meet being a woman. The probability distribution in Equation (2.1) is the Bernoulli distribution.

FIGURE 2.2 Cumulative Distribution and Probability Density Functions of Commuting Time



(a) Cumulative distribution function of commuting time



(b) Probability density function of commuting time

Figure 2.2a shows the cumulative probability distribution (or c.d.f.) of commuting times. The probability that a commuting time is less than 15 minutes is 0.20 (or 20%), and the probability that it is less than 20 minutes is 0.78 (78%). Figure 2.2b shows the probability density function (or p.d.f.) of commuting times. Probabilities are given by areas under the p.d.f. The probability that a commuting time is between 15 and 20 minutes is 0.58 (58%) and is given by the area under the curve between 15 and 20 minutes.

2.3 Two Random Variables

Most of the interesting questions in economics involve two or more variables. Are college graduates more likely to have a job than nongraduates? How does the distribution of income for women compare to that for men? These questions concern the distribution of two random variables, considered together (education and employment status in the first example, income and gender in the second). Answering such questions requires an understanding of the concepts of joint, marginal, and conditional probability distributions.

Joint and Marginal Distributions

Joint distribution. The **joint probability distribution** of two discrete random variables, say X and Y, is the probability that the random variables simultaneously take on certain values, say x and y. The probabilities of all possible (x, y) combinations sum to 1. The joint probability distribution can be written as the function $\Pr(X = x, Y = y)$.

For example, weather conditions—whether or not it is raining—affect the commuting time of the student commuter in Section 2.1. Let Y be a binary random variable that equals 1 if the commute is short (less than 20 minutes) and equals 0 otherwise and let X be a binary random variable that equals 0 if it is raining and 1 if not. Between these two random variables, there are four possible outcomes: it rains and the commute is long $(X = 0, Y = 0)$; rain and short commute $(X = 0, Y = 1)$; no rain and long commute $(X = 1, Y = 0)$; and no rain and short commute $(X = 1, Y = 1)$. The joint probability distribution is the frequency with which each of these four outcomes occurs over many repeated commutes.

An example of a joint distribution of these two variables is given in Table 2.2. According to this distribution, over many commutes, 15% of the days have rain and a long commute $(X = 0, Y = 0)$; that is, the probability of a long, rainy commute is 15%, or $\Pr(X = 0, Y = 0) = 0.15$. Also, $\Pr(X = 0, Y = 1) = 0.15$, $\Pr(X = 1,$

TABLE 2.2 Joint Distribution of Weather Conditions and Commuting Times

	Rain ($X = 0$)	No Rain ($X = 1$)	Total
Long commute ($Y = 0$)	0.15	0.07	0.22
Short commute ($Y = 1$)	0.15	0.63	0.78
Total	0.30	0.70	1.00

TABLE 2.3 Joint and Conditional Distributions of Computer Crashes (M) and Computer Age (A)

A. Joint Distribution

	$M = 0$	$M = 1$	$M = 2$	$M = 3$	$M = 4$	**Total**
Old computer ($A = 0$)	0.35	0.065	0.05	0.025	0.01	0.50
New computer ($A = 1$)	0.45	0.035	0.01	0.005	0.00	0.50
Total	0.80	0.10	0.06	0.03	0.01	1.00

B. Conditional Distributions of *M* given *A*

	$M = 0$	$M = 1$	$M = 2$	$M = 3$	$M = 4$	**Total**
$\Pr(M \mid A = 0)$	0.70	0.13	0.10	0.05	0.02	1.00
$\Pr(M \mid A = 1)$	0.90	0.07	0.02	0.01	0.00	1.00

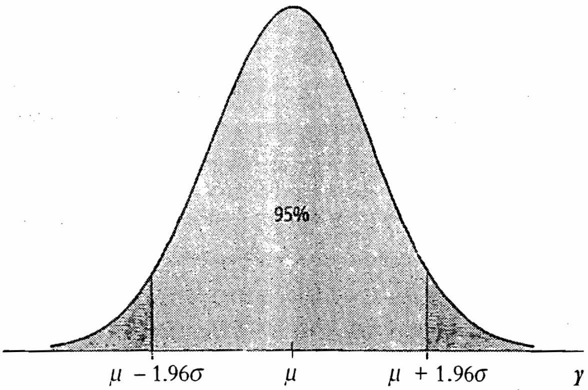
For example, the conditional probability of a long commute given that it is rainy is $\Pr(Y = 0 \mid X = 0) = \Pr(X = 0, Y = 0)/\Pr(X = 0) = 0.15/0.30 = 0.50$.

As a second example, consider a modification of the crashing computer example. Suppose you use a computer in the library to type your term paper and the librarian randomly assigns you a computer from those available, half of which are new and half of which are old. Because you are randomly assigned to a computer, the age of the computer you use, A ($= 1$ if the computer is new, $= 0$ if it is old), is a random variable. Suppose the joint distribution of the random variables M and A is given in Part A of Table 2.3. Then the conditional distribution of computer crashes, given the age of the computer, is given in Part B of the table. For example, the joint probability $M = 0$ and $A = 0$ is 0.35; because half the computers are old, the conditional probability of no crashes, given that you are using an old computer, is $\Pr(M = 0 \mid A = 0) = \Pr(M = 0,\ A = 0)/\Pr(A = 0) = 0.35/0.50 = 0.70$, or 70%. In contrast, the conditional probability of no crashes given that you are assigned a new computer is 90%. According to the conditional distributions in Part B of Table 2.3, the newer computers are less likely to crash than the old ones; for example, the probability of three crashes is 5% with an old computer but 1% with a new computer.

Conditional expectation. The **conditional expectation** of Y given X, also called the **conditional mean** of Y given X, is the mean of the conditional distribution of Y given X. That is, the conditional expectation is the expected value of Y, computed

FIGURE 2.5 The Normal Probability Density

The normal probability density function with mean μ and variance σ^2 is a bell-shaped curve, centered at μ. The area under the normal p.d.f. between $\mu - 1.96\sigma$ and $\mu + 1.96\sigma$ is 0.95. The normal distribution is denoted $N(\mu, \sigma^2)$.



2.4 The Normal, Chi-Squared, Student *t*, and *F* Distributions

The probability distributions most often encountered in econometrics are the normal, chi-squared, Student t, and F distributions.

The Normal Distribution

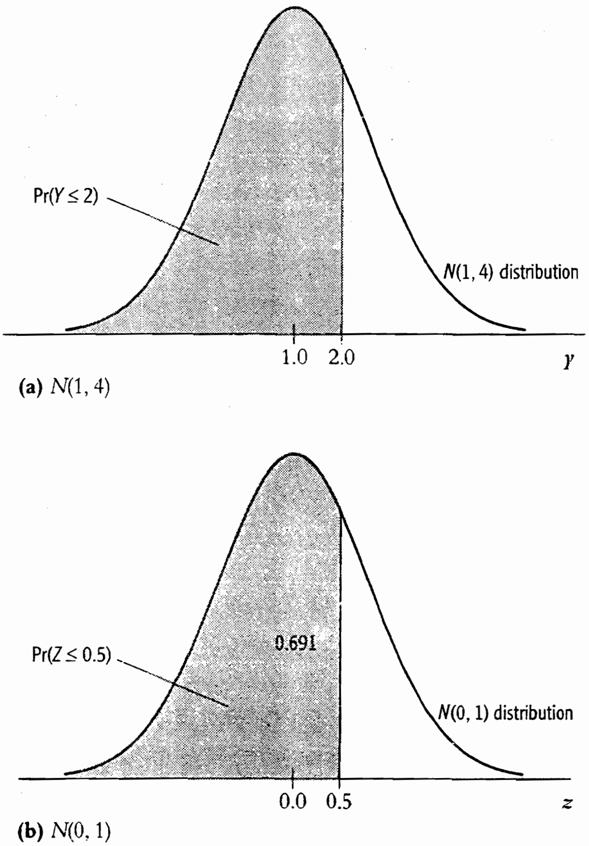
A continuous random variable with a **normal distribution** has the familiar bell-shaped probability density shown in Figure 2.5. The function defining the normal probability density is given in Appendix 17.1. As Figure 2.5 shows, the normal density with mean μ and variance σ^2 is symmetric around its mean and has 95% of its probability between $\mu - 1.96\sigma$ and $\mu + 1.96\sigma$.

Some special notation and terminology have been developed for the normal distribution. The normal distribution with mean μ and variance σ^2 is expressed concisely as "$N(\mu, \sigma^2)$." The **standard normal distribution** is the normal distribution with mean $\mu = 0$ and variance $\sigma^2 = 1$ and is denoted $N(0, 1)$. Random variables that have a $N(0, 1)$ distribution are often denoted Z, and the standard normal cumulative distribution function is denoted by the Greek letter Φ; accordingly, $\Pr(Z \leq c) = \Phi(c)$, where c is a constant. Values of the standard normal cumulative distribution function are tabulated in Appendix Table 1.

To look up probabilities for a normal variable with a general mean and variance, we must **standardize the variable** by first subtracting the mean, then by dividing

FIGURE 2.6 Calculating the Probability That $Y \leq 2$ When Y Is Distributed $N(1, 4)$

To calculate $\Pr(Y \leq 2)$, standardize Y, then use the standard normal distribution table. Y is standardized by subtracting its mean ($\mu = 1$) and dividing by its standard deviation ($\sigma = 2$). The probability that $Y \leq 2$ is shown in Figure 2.6a, and the corresponding probability after standardizing Y is shown in Figure 2.6b. Because the standardized random variable, $(Y - 1)/2$, is a standard normal (Z) random variable, $\Pr(Y \leq 2) = \Pr\left(\frac{Y-1}{2} \leq \frac{2-1}{2}\right) = \Pr(Z \leq 0.5)$. From Appendix Table 1, $\Pr(Z \leq 0.5) = \Phi(0.5) = 0.691$.

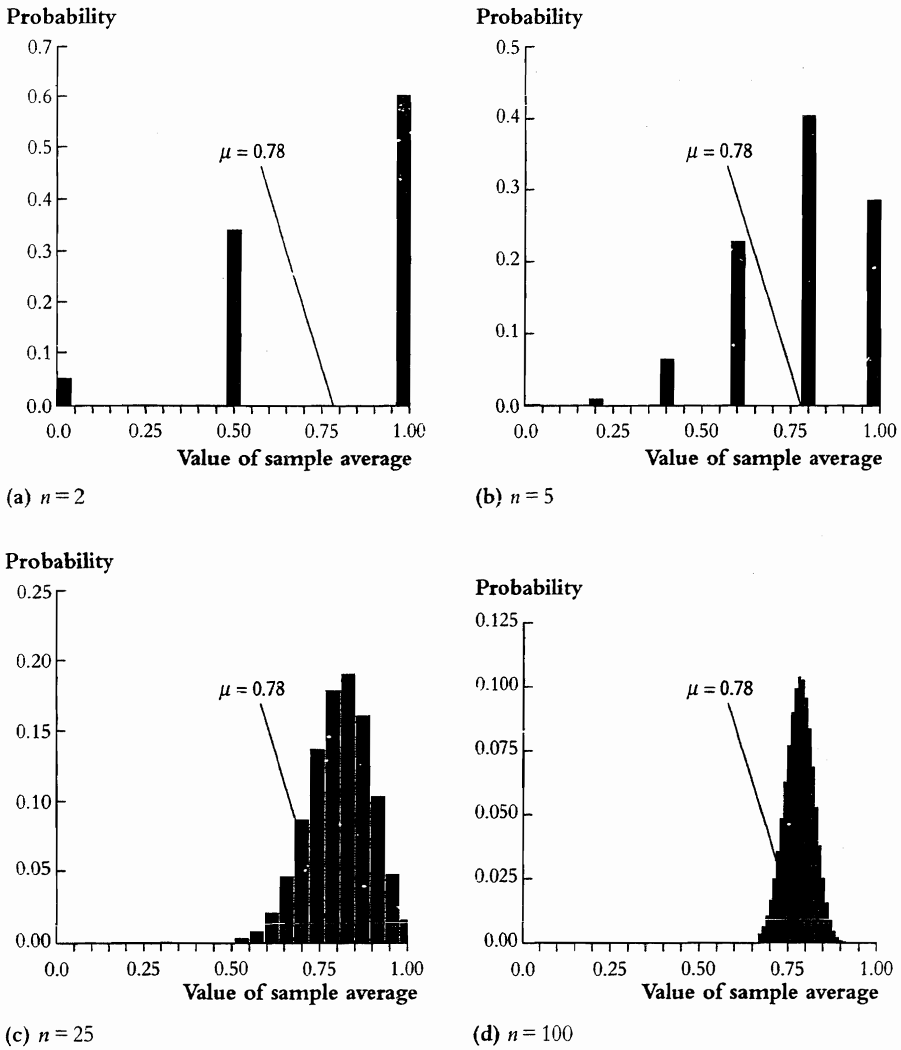


The multivariate normal distribution. The normal distribution can be generalized to describe the joint distribution of a set of random variables. In this case, the distribution is called the **multivariate normal distribution**, or, if only two variables are being considered, the **bivariate normal distribution**. The formula for the bivariate normal p.d.f. is given in Appendix 17.1, and the formula for the general multivariate normal p.d.f. is given in Appendix 18.1.

The multivariate normal distribution has four important properties. If X and Y have a bivariate normal distribution with covariance σ_{XY} and if a and b are two constants, then $aX + bY$ has the normal distribution:

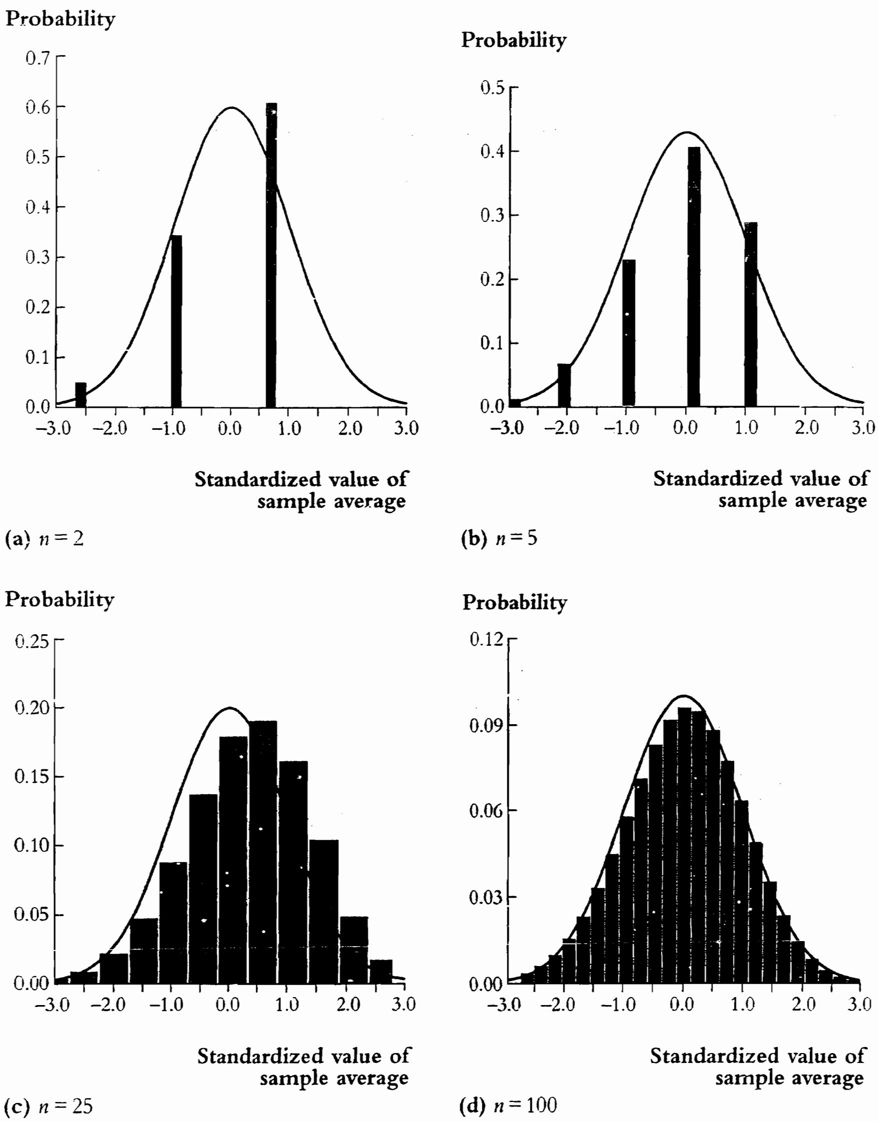
$$aX + bY \text{ is distributed } N(a\mu_X + b\mu_Y,\ a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\sigma_{XY}) \quad (X, Y \text{ bivariate normal}) \tag{2.42}$$

FIGURE 2.8 Sampling Distribution of the Sample Average of n Bernoulli Random Variables



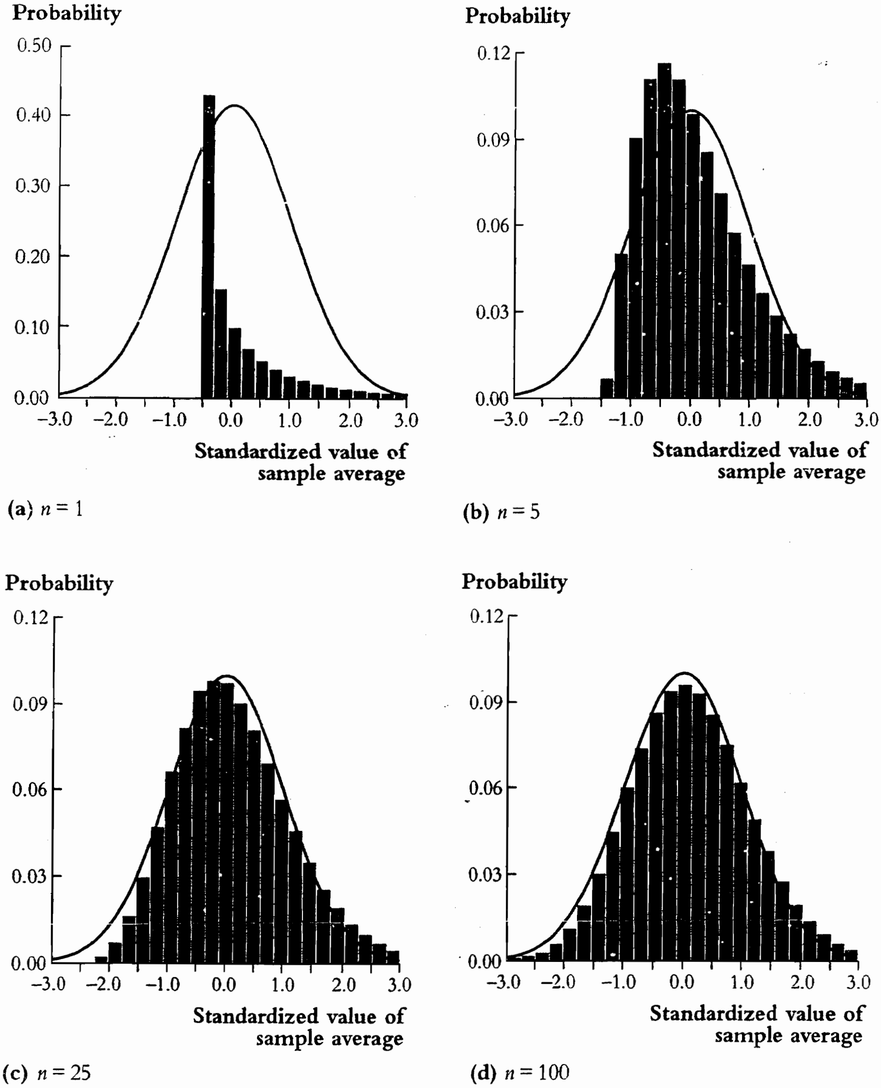
The distributions are the sampling distributions of $\overline{Y}$, the sample average of n independent Bernoulli random variables with $p = \Pr(Y_i = 1) = 0.78$ (the probability of a short commute is 78%). The variance of the sampling distribution of $\overline{Y}$ decreases as n gets larger, so the sampling distribution becomes more tightly concentrated around its mean $\mu = 0.78$ as the sample size n increases.

FIGURE 2.9 Distribution of the Standardized Sample Average of n Bernoulli Random Variables with $p = 0.78$



The sampling distribution of $\overline{Y}$ in Figure 2.8 is plotted here after standardizing $\overline{Y}$. This plot centers the distributions in Figure 2.8 and magnifies the scale on the horizontal axis by a factor of $\sqrt{n}$. When the sample size is large, the sampling distributions are increasingly well approximated by the normal distribution (the solid line), as predicted by the central limit theorem. The normal distribution is scaled so that the height of the distributions is approximately the same in all figures.

FIGURE 2.10 Distribution of the Standardized Sample Average of *n* Draws from a Skewed Distribution



The figures show the sampling distribution of the standardized sample average of n draws from the skewed (asymmetric) population distribution shown in Figure 2.10a. When n is small ($n = 5$), the sampling distribution, like the population distribution, is skewed. But when n is large ($n = 100$), the sampling distribution is well approximated by a standard normal distribution (solid line), as predicted by the central limit theorem. The normal distribution is scaled so that the height of the distributions is approximately the same in all figures.