APPENDIX B

CHARACTERISTICS OF PROBABILITY DISTRIBUTIONS

Although a PMF (PDF) indicates the values taken by a random variable (r.v.) and their associated probabilities, often we are not interested in the entire PMF. Thus, in the PMF of Example A.13 we may not want the individual probabilities of obtaining no heads, one head, or two heads. Rather, we may wish to find out the *average number* of heads obtained when tossing a coin several times. In other words, we may be interested in some summary **characteristics,** or more technically, the **moments** of a probability distribution. Two of the most commonly used summary measures or moments are the *expected value* (called the *first moment* of the probability distribution) and the *variance* (called the *second moment* of the probability distribution). On occasion, we will need higher moments of probability distributions, which we will discuss as we progress.

B.1 EXPECTED VALUE: A MEASURE OF CENTRAL TENDENCY

The **expected value** of a discrete r.v. X, denoted by the symbol $E(X)$ (read as E of X), is defined as follows:

$$E(X) = \sum_X xf(X) \tag{B.1}$$

where $f(X)$ is the PMF of X and where $\sum_X$ means the sum over all values of X.[1]

Verbally, the expected value of a random variable is the *weighted average* of its possible values, with the probabilities of these values [i.e., $f(X)$] serving as the *weights.* Equivalently, *it is the sum of products of the values taken by the r.v. and their corresponding probabilities.* The expected value of an r.v. is also known as its *average*

[1]The expected value of a continuous r.v. is defined similarly, with the summation symbol being replaced by the integral symbol. That is: $E(X) = \int xf(x)dx$, where the integral is over all the values of X.

TABLE B-1 THE EXPECTED VALUE OF A RANDOM VARIABLE X, THE NUMBER SHOWN ON A DIE

Number shown (1) X	Probability (2) $f(X)$	Number × Probability (3) $Xf(X)$
1	1/6	1/6
2	1/6	2/6
3	1/6	3/6
4	1/6	4/6
5	1/6	5/6
6	1/6	6/6
		$E(X) = 21/6 = 3.5$

or *mean* value, although, more correctly, it is called the **population mean value** for reasons to be discussed shortly.

Example B.1.

Suppose we roll a die numbered 1 through 6 several times. What is the expected value of the number shown? As given previously (see Example A.6), we have the situation shown in Table B-1.

Applying the definition of the expected value given in Eq. (B.1), we see that the expected value is 3.5.

Is it strange that we obtained this value, since the r.v. here is discrete and can take only one of the six values 1 through 6? The expected, or average, value of 3.5 in this example means that if we were to roll the die several times, then on the *average*, we would obtain the number 3.5, which is between 3 and 4. If, in a contest, someone were to give you as many dollars as the number shown on the die, then in several rolls of the die you would anticipate receiving on the average $3.50 per roll of the die.

Geometrically, the expected value of the preceding example is shown in Figure B-1.

Example B.2.

In the PC/printer sales example (Example A.17), what is the expected value of the number of PCs sold? This can be obtained easily from Table A-4 by multiplying the values of X (PCs sold) by their associated probabilities (i.e., $f[X]$) and summing the product. Thus,

$$E(X) = 0(0.08) + 1(0.12) + 2(0.24) + 3(0.24) + 4(0.32) = 2.60$$

That is, the average number of PCs sold per day is 2.60. Keep in mind that this is an average. On any given day the number of PCs sold will be any one of the numbers between 0 and 4.

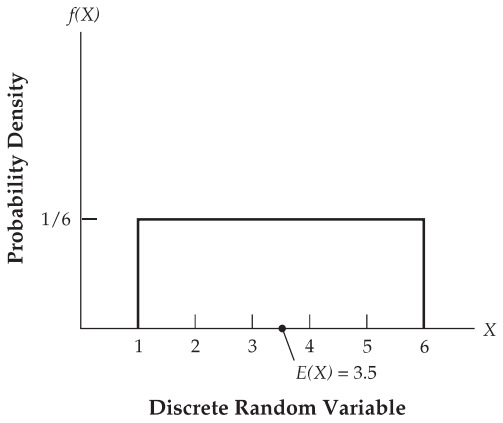

FIGURE B-1 The expected value, $E(X)$, of a discrete random variable (Example B.1)

You can easily verify that $E(Y) = 2.35$; that is, the average number of printers sold is 2.35.

Properties of Expected Value

The following properties of the expected value will prove very useful in the main chapters of the text:

1. The expected value of a constant is that constant itself. Thus, if b is a constant,

$$E(b) = b \tag{B.2}$$

   For example, if $b = 2$, $E(2) = 2$.
2. The expectation of the sum of two random variables is equal to the sum of the expectations of those random variables. Thus, for the random variables X and Y:[2]

$$E(X + Y) = E(X) + E(Y) \tag{B.3}$$

3. However,

$$E(X/Y) \neq \frac{E(X)}{E(Y)} \tag{B.4}$$

[2]This property can be generalized to more than two random variables. Thus, $E(X + Y + W + Z) = E(X) + E(Y) + E(W) + E(Z)$.

That is, the expected value of the ratio of two random variables is not equal to the ratio of the expected values of those random variables.

4. Also, in general,

$$E(XY) \neq E(X)E(Y) \tag{B.5}$$

That is, in general, the expected value of the product of two random variables is not equal to the product of the expectations of those random variables. However, there is an exception to the rule. If X and Y are *independent random variables*, then it is true that

$$E(XY) = E(X)E(Y) \tag{B.6}$$

Recall that X and Y are said to be independent if and only if $f(X, Y) = f(X)f(Y)$, for all values of X and Y, that is, when the joint PMF (PDF) is equal to the product of the individual PMFs (PDFs) of the two random variables for all values of the variables.

5. $$E(X^2) \neq [E(X)]^2 \tag{B.7}$$

That is, the expected value of the square of X (or any random variable) is *not* equal to the square of the expected value of X.

6. If a is a constant, then

$$E(aX) = aE(X) \tag{B.8}$$

That is, the expectation of a constant times an r.v. is equal to the constant times the expectation of the r.v.

7. If a and b are constants, then

$$\begin{aligned} E(aX + b) &= aE(X) + E(b) \\ &= aE(X) + b \end{aligned} \tag{B.9}$$

In deriving result (7), we use properties (1), (2), and (6). Thus,

$$E(4X + 7) = 4E(X) + E(7) = 4E(X) + 7$$

From Eq. (B.9) we see that E is a *linear operator*, which is also evident from Eq. (B.4).

Expected Value of Multivariate Probability Distributions

The concept of the expected value of a random variable can be extended easily to multivariate PMF or PDF. In the bivariate PMF, it can be shown that

$$E(XY) = \sum_x \sum_y xyf(X, Y) \tag{B.10}$$

That is, we take each pair of X and Y values, multiply them by their joint probability, and sum over all the values of X and Y.

Example B.3.

Continuing with our PC/printer sales example, and applying Eq. (B.10), we get

$$\begin{aligned} E(XY) &= (1)(1)(0.05) + (1)(2)(0.06) + (1)(3)(0.02) + (1)(4)(0.01) \cdots \\ &\quad + (4)(1)(0.01) + 4(2)(0.01) + 4(3)(0.01) + (4)(3)(0.05) + (4)(4)(0.15) \\ &= 7.06 \end{aligned}$$

which is the expected value of the product of the two random variables.

Recall that if two variables are independent, the expected value of their product is equal to the product of their individual expected values; that is, $E(XY) = E(X)E(Y)$. Is this the case in our illustrative example? As we saw in Example B.2, $E(X) = 2.60$ and $E(Y) = 2.35$. Therefore, $E(X)E(Y) = (2.60)(2.35) = 6.11 \neq E(XY) = 7.06$, showing that the two variables are not independent.

In passing, note that the formula for the expected value of the product of two random variables given in Eq. (B.10) is for two discrete random variables. In the case of two continuous random variables, in Eq. (B.10) we would replace the double summation sign by the double integral sign.

B.2 VARIANCE: A MEASURE OF DISPERSION

The expected value of an r.v. simply gives its *center of gravity,* but it does not indicate how the individual values are spread, dispersed, or distributed around this mean value. The most popular numerical measure of this spread is called the **variance,** which is defined as follows.

Let X be an r.v. and $E(X)$ be its expected value, which for notational simplicity may be denoted by μ_x (where μ is the Greek letter mu). Then the variance of X is defined as

$$\text{var}(X) = \sigma_x^2 = E(X - \mu_x)^2 \tag{B.11}$$

where $\mu_x = E(X)$ and where the Greek letter σ_x^2 (sigma squared) is the commonly used symbol for the variance. As Equation (B.11) shows, the variance of X (or any r.v.) is simply the expected value of the squared difference between an individual X value and its expected or mean value. The variance thus defined shows how the individual X values are spread or distributed around its expected, or mean, value. If all X values are precisely equal to $E(X)$, the variance will be zero, whereas if they are widely spread around the expected value, it will be relatively large, as shown in Figure B-2. Notice that the variance cannot be a negative number. (Why?)

The positive square root of σ_x^2, σ_x, is known as the **standard deviation (s.d.).** Equation (B.11) is the definition of variance. To compute the variance, we use the following formula:

$$\text{var}(X) = \sum_X (X - \mu_x)^2 f(X) \tag{B.12}$$

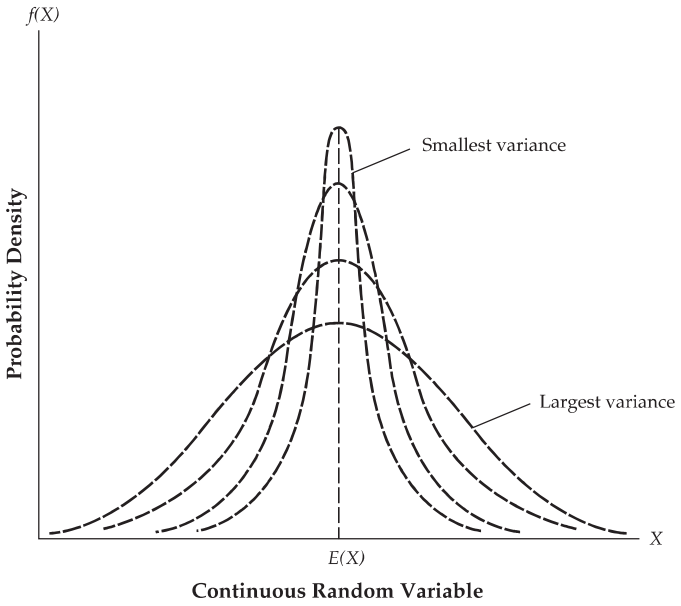

FIGURE B-2 Hypothetical PDFs of continuous random variables all with the same expected value

if X is a discrete r.v. In case of a continuous random variable, we replace the summation symbol by the integral symbol.

As Equation (B.12) shows, to compute the variance of a discrete r.v., we subtract the expected value of the variable from a given value of the variable, square the difference, and multiply the squared difference by the probability associated with that X value. We do this for each value assumed by the X variable and sum the products thus obtained. An example follows.

Example B.4.

We continue with Example B.1. There we showed that the expected value of the number in the repeated roll of a die is 3.5. To compute the variance for that problem, we set up Table B-2.

Thus, the variance of this example is 2.9167. Taking the positive square root of this value, we obtain a standard deviation (s.d.) of 1.7078.

Properties of Variance

The variance as defined earlier has the following properties, which we will find useful in our discussion of econometrics in the main chapters of the text.

1. The variance of a constant is zero. By definition, a constant has no variability.

TABLE B-2 THE VARIANCE OF A RANDOM VARIABLE X, THE NUMBER SHOWN ON A DIE

Number Shown	Probability	
X	$f(X)$	$(X-\mu_X)^2 f(X)$
1	1/6	$(1-3.5)^2$ (1/6)
2	1/6	$(2-3.5)^2$ (1/6)
3	1/6	$(3-3.5)^2$ (1/6)
4	1/6	$(4-3.5)^2$ (1/6)
5	1/6	$(5-3.5)^2$ (1/6)
6	1/6	$(6-3.5)^2$ (1/6)
		Sum = 2.9167

2. If X and Y are two *independent* random variables, then

$$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y)$$

and

$$\text{var}(X - Y) = \text{var}(X) + \text{var}(Y) \quad \textbf{(B.13)}$$

That is, the variance of the sum or difference of two independent random variables is equal to the sum of their individual variances.

3. If b is a constant, then

$$\text{var}(X + b) = \text{var}(X) \quad \textbf{(B.14)}$$

That is, adding a constant number to (the values of) a variable does not change the variance of that variable. Thus, $\text{var}(X + 7) = \text{var}(X)$.

4. If a is constant, then

$$\text{var}(aX) = a^2\text{var}(X) \quad \textbf{(B.15)}$$

That is, the variance of a constant times a variable is equal to the square of that constant times the variance of that variable. Thus, $\text{var}(5X) = 25\ \text{var}(X)$.

5. If a and b are constant, then

$$\text{var}(aX + b) = a^2\text{var}(X) \quad \textbf{(B.16)}$$

which follows from properties (3) and (4). Thus,

$$\text{var}(5X + 9) = 25\ \text{var}(X)$$

6. If X and Y are *independent* random variables and a and b are constants, then

$$\text{var}(aX + bY) = a^2\text{var}(X) + b^2\text{var}(Y) \quad \textbf{(B.17)}$$

This property follows from the previous properties. Thus,

$$\text{var}(3X + 5Y) = 9\,\text{var}(X) + 25\,\text{var}(Y)$$

7. For computational convenience, the variance formula Eq. (B.11) can also be written as

$$\text{var}(X) = E(X^2) - [E(X)]^2 \tag{B.18}$$

which says that the variance of X is equal to the expected value of X squared minus the square of the expected value of X.[3] Note that

$$E(X^2) = \sum_x x^2 f(X) \tag{B.19}$$

for a discrete r.v. For a continuous r.v., replace the summation sign with the integral sign.

The proofs of the various expressions above can be obtained from the basic definition of variance (see the optional exercises given at the end of this appendix).

Chebyshev's Inequality

How adequate are the expected value and variance of a random variable to describe a PMF or PDF of such a random variable? That is, knowing just these two summary numbers of a random variable, say X, can we compute the probability that X lies in a certain range? In a remarkable theorem, known as *Chebyshev's inequality*, the Russian mathematician Pafnuty Lvovich Chebyshev (1821–1894) showed that that is indeed possible.

Specifically, if X is a random variable with mean μ_x and a variance of σ_x^2, then for any positive constant c the probability that X lies *inside* the interval $[\mu_x - c\sigma_x, \mu_x + c\sigma_x]$ is at least $1 - \frac{1}{c^2}$, that is

$$P[|X - \mu_x| \leq c\sigma_x] \leq 1 - \frac{1}{c^2} \tag{B.20}$$

where the symbol $|\ |$ means the absolute value of.[4]

[3]The proof is as follows:

$$\begin{aligned} E(X - \mu_x)^2 &= E\left(X^2 - 2X\mu_x + \mu_x^2\right) \\ &= E(X^2) - 2\mu_x E(X) + E\left(\mu_x^2\right) \\ &= E(X^2) - 2\mu_x^2 + \mu_x^2 = E(X^2) - \mu_x^2 \end{aligned}$$

Keep in mind that μ_x is a constant.

[4]The inequality works quite well if $c > 1$.

In words this inequality states that at least the fraction $(1 - \frac{1}{c^2})$ of the total probability of X lies within c standard deviations of its mean or expected value. Put differently, the probability that a random variable deviates from its mean value by more than c standard deviations is less than or at the most equal to $1/c^2$.

What is remarkable about this inequality is that we do not need to know the actual PDF or PMF of a random variable. Of course, if we know the actual PDF or PMF, probabilities such as Eq. (B.20) can be computed easily, as we will show when we consider some specific probability distributions in Appendix C.

Example B.5. Illustration of Chebyshev's Inequality

The average number of donuts sold in a donut shop between 8 a.m. and 9 a.m. is 100 with a variance of 25. What is the probability that on a given day the number of donuts sold between 8 a.m. and 9 a.m. is between 90 and 110?

By Chebyshev's inequality, we have:

$$P[|X - \mu_x| \leq c\sigma_x] = 1 - \frac{1}{c^2}$$
$$P[|X - 100| \leq 5c] = 1 - \frac{1}{c^2} \tag{B.21}$$

Since $(110 - 100) = (100 - 90) = 10$, we see that $5c = 10$. Therefore, $c = 2$. It therefore follows that $(1 - \frac{1}{2^2}) = \frac{3}{4} = 0.75$. That is, the probability that between 90 and 110 donuts are sold between 8 a.m. and 9 a.m. is at least 75 percent. By the same token, the probability that the number of donuts sold between 8 a.m. and 9 a.m. exceeds 110 or is less than 90 is 25 percent.

Coefficient of Variation

Before moving on, note that since the standard deviation (or variance) depends on the units of measurement, it may be difficult to compare two or more standard deviations if they are expressed in different units of measurement. To get around this difficulty, use the **coefficient of variation (*V*),** a measure of *relative variation*, which is defined as follows:

$$V = \frac{\sigma_X}{\mu_X} \cdot 100 \tag{B.22}$$

Verbally, the V is the ratio of the standard deviation of a random variable X to its mean value multiplied by 100. Since the standard deviation and the mean value of a random variable are measured in the same units of measurement, V is unitless; that is, it is a pure number. We can therefore compare the V values of two or more random variables directly.

Example B.6.

An instructor teaches two sections of an introductory econometrics class with 15 students in each class. On the midterm examination, class A scored

an average of 83 points with a standard deviation of 10, and class B scored an average 88 points with a standard deviation of 16. Which class performed relatively better? If we use V as defined in Eq. (B.22), we get:

$$V_A = \frac{10}{83} \cdot 100 = 12.048 \quad \textit{and} \quad V_B = \frac{16}{88} \cdot 100 = 18.181$$

Since the relative variability of class A is lower, we can say that class A did relatively better than class B.

B.3 COVARIANCE

The expected value and variance are the two most frequently used summary measures of a univariate PMF (or PDF). The former gives us the center of gravity, and the latter tells us how the individual values are distributed around the center of gravity. But once we go beyond the univariate probability distributions (e.g., the PMF of Example B.2), we need to consider, in addition to the mean and variance of each variable, some additional characteristics of multivariate PFs, such as the ***covariance*** and ***correlation,*** which we will now discuss.

Let X and Y be two random variables with means $E(X) = \mu_x$ and $E(Y) = \mu_y$. Then the covariance (cov) between the two variables is defined as

$$\begin{aligned} \text{cov}(X, Y) &= E[(X - \mu_x)(Y - \mu_y)] \\ &= E(XY) - \mu_x\mu_y \end{aligned} \tag{B.23}$$

As Equation (B.23) shows, a **covariance** is a special kind of expected value and is a measure of how two variables vary or move together (i.e., co-vary), as shown in Example B.7, which follows. In words, Eq. (B.23) states that to find the covariance between two variables, we must express the value of each variable as a deviation from its mean value and take the expected value of the product. How this is done in practice follows.

The covariance between two random variables can be *positive*, *negative*, or *zero.* If two random variables move in the same direction (i.e., if they both increase) as in Example B.7 below, then the covariance will be positive, whereas if they move in the opposite direction (i.e., if one increases and the other decreases), the covariance will be negative. If, however, the covariance between the two variables is zero, it means that there is no (linear) relationship between the two variables.

To compute the covariance as defined in Eq. (B.23), we use the following formula, assuming X and Y are discrete random variables:

$$\begin{aligned} \text{cov}(X, Y) &= \sum_x \sum_y (X - \mu_X)(Y - \mu_y) f(X, Y) \\ &= \sum_x \sum_y XY f(X, Y) - \mu_X\mu_y \\ &= E(XY) - \mu_X\mu_y \end{aligned} \tag{B.24}$$

where $E(XY)$ is computed from Eq. (B.10).

Note the double summation sign in this expression because the covariance requires the summation of both variables over the range of their values. Using the integral notation of calculus, a similar formula can be devised to compute the covariance of two continuous random variables.

Example B.7.

Once again, return to our PC/printer sales example. To find out the covariance between computer sales (X) and printer sales (Y), we use formula (B.24). We have already computed the first term on the right-hand side of this equation in Example (B.3), which is 7.06. We have already found that $\mu_x = 2.60$ and $\mu_y = 2.35$. Therefore, the covariance in this example is

$$\text{cov}(X, Y) = 7.06 - (2.60)(2.35) = 0.95$$

which shows that PC sales and printer sales are positively related.

Properties of Covariance

The covariance as defined earlier has the following properties, which we will find quite useful in regression analysis in the main chapters of the text.

1. If X and Y are *independent* random variables, their covariance is zero. This is easy to verify. Recall that if two random variables are independent,

   $$E(XY) = E(X)E(Y) = \mu_x\mu_y$$

   Substituting this expression into Eq. (B.23), we see at once that the covariance of two *independent* random variables is zero.

2. $$\text{cov}(a + bX, c + dY) = bd\,\text{cov}(X, Y) \qquad \textbf{(B.25)}$$

   where a, b, c, and d are constants.

3. $$\text{cov}(X, X) = \text{var}(X) \qquad \textbf{(B.26)}$$

   That is, the covariance of a variable with itself is simply its variance, which can be verified from the definitions of variance and covariance given previously. Obviously, then, $\text{cov}(Y, Y) = \text{var}(Y)$.

4. If X and Y are two random variables but are not necessarily independent, then the variance formulas given in Eq. (B.13) need to be modified as follows:

   $$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) + 2\,\text{cov}(X, Y) \qquad \textbf{(B.27)}$$

   $$\text{var}(X - Y) = \text{var}(X) + \text{var}(Y) - 2\,\text{cov}(X, Y) \qquad \textbf{(B.28)}$$

Of course, if the two variables are independent, formulas (B.27) and (B.28) will coincide with Eq. (B.13).

B.4 CORRELATION COEFFICIENT

In the PC/printer sales example just considered we found that the covariance between PC sales and computer sales was 0.95, which suggests that the two variables are positively related. But the computed number of 0.95 does not give any idea of how strongly the two variables are positively related because the covariance is unbounded (i.e., $-\infty < \text{cov}[X, Y] < \infty$). We can find out how strongly any two variables are related in terms of what is known as the **(population) coefficient of correlation,** which is defined as follows:

$$\rho = \frac{\text{cov}(X, Y)}{\sigma_x \sigma_y} \tag{B.29}$$

where ρ (rho) denotes the coefficient of correlation.

As is clear from Equation (B.29), the **correlation** between two random variables X and Y is simply the ratio of the covariance between the two variables divided by their respective standard deviations. The correlation coefficient thus defined is a measure of *linear* association between two variables, that is, how strongly the two variables are linearly related.

Properties of Correlation Coefficient

The correlation coefficient just defined has the following properties:

1. Like the covariance, the correlation coefficient can be positive or negative. It is positive if the covariance is positive and negative if the covariance is negative. In short, it has the same sign as the covariance.
2. The correlation coefficient is a measure of *linear relationship* between two variables.
3. The correlation coefficient always lies between −1 and +1. Symbolically,

$$-1 \leq \rho \leq 1 \tag{B.30}$$

   If the correlation coefficient is +1, it means that the two variables are perfectly positively linearly related (as in $Y = B_1 + B_2X$), whereas if the correlation coefficient is −1, it means they are perfectly negatively linearly related. Typically, ρ lies between these limits.
4. The correlation coefficient is a *pure number*; that is, it is devoid of units of measurement. On the other hand, other characteristics of probability distributions, such as the expected value, variance, and covariance, depend on the units in which the original variables are measured.
5. If two variables are (statistically) independent, their covariance is zero. Therefore, the correlation coefficient will be zero. The converse, however, is not true. That is, if the correlation coefficient between two variables is zero, it does not mean that the two variables are independent. This is because the correlation coefficient is a measure of *linear association* or *linear relationship* between two variables, as noted previously. For example, if $Y = X^2$, the correlation between the two variables may be zero, but by no means are the two variables independent. Here Y is a *nonlinear* function of X.

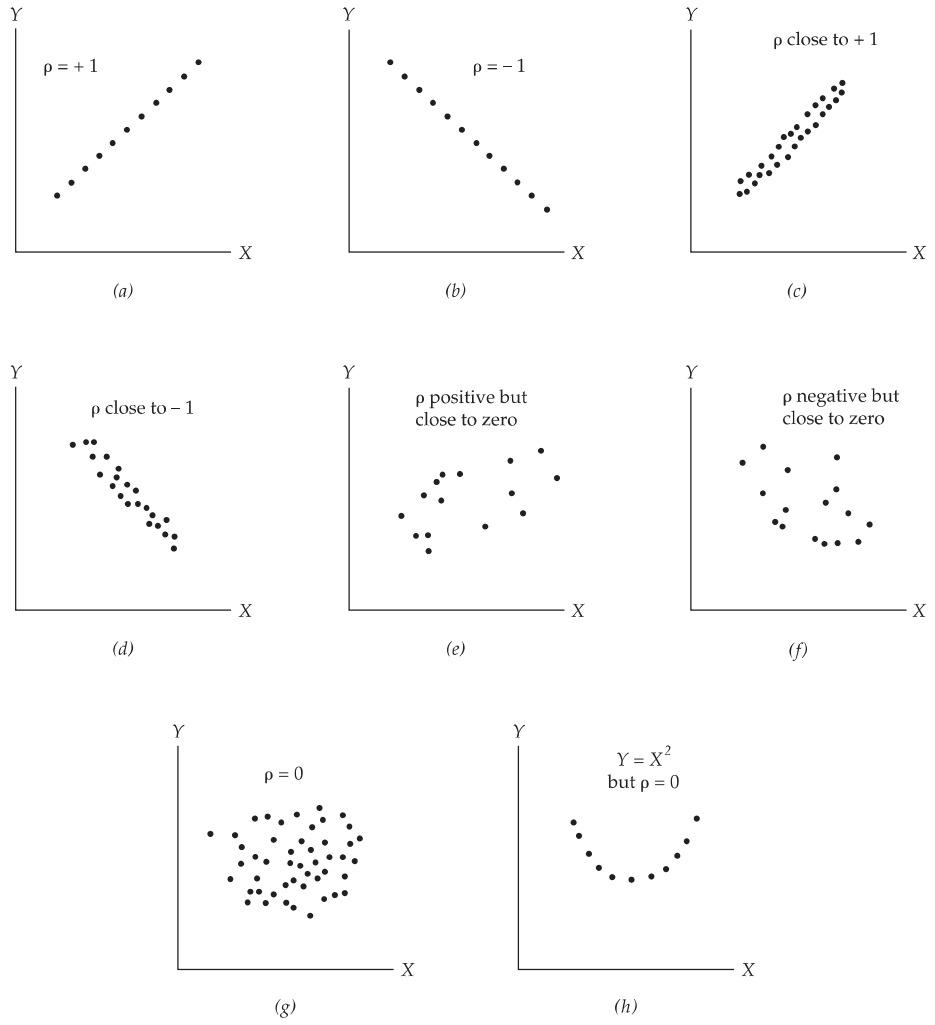

FIGURE B-3 Some typical patterns of the correlation coefficient, ρ

6. Correlation does not necessarily imply causality. If one finds a positive correlation between lung cancer and smoking, it does not necessarily mean that smoking causes lung cancer.

Figure B-3 gives some typical patterns of correlation coefficients.

Example B.8.

Let us continue with the PC/printer sales example. We have already seen that the covariance between the two variables is 0.95. From the data given in

Table A-4, we can easily verify that $\sigma_x = 1.2649$ and $\sigma_y = 1.4124$. Then, using formula (B.29), we obtain

$$\rho = \frac{0.95}{(1.2649)(1.4124)} = 0.5317$$

Thus, the two variables are positively correlated, although the value of the correlation coefficient is rather moderate. This probably is not surprising, for not everyone purchasing a PC buys a printer.

The use of the correlation coefficient in the regression context is discussed in Chapter 3.

Incidentally, Eq. (B.29) can also be written as:

$$\text{Cov}(X, Y) = \rho\sigma_x\sigma_y \tag{B.31}$$

That is, the covariance between two variables is equal to the coefficient of correlation between the two times the product of the standard deviations of the two.

Variances of Correlated Variables

In Eq. (B.27) and Eq. (B.28) we gave formulas for the variance of variables that are not necessarily independent. Knowing the relationship between covariance and correlation, we can express these formulas alternatively as follows:

$$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) + 2\rho\sigma_x\sigma_y \tag{B.32}$$

$$\text{var}(X - Y) = \text{var}(X) + \text{var}(Y) - 2\rho\sigma_x\sigma_y \tag{B.33}$$

Of course, if the correlation between two random variables is zero, then $\text{var}(X + Y)$ and $\text{var}(X - Y)$ are both equal to $\text{var}(X) + \text{var}(Y)$, as we saw before.

As an exercise, you can find the variance of $(X + Y)$ of our PC/printer example.

B.5 CONDITIONAL EXPECTATION

Another statistical concept that is especially important in regression analysis is the concept of **conditional expectation,** which is different from the expectation of an r.v. considered previously, which may be called the **unconditional expectation.** The difference between the two concepts of expectations can be explained as follows.

Return to our PC/printer sales example. In this example X is the number of PCs sold per day (ranging from 0 to 4) and Y is the number of printers sold per day (ranging from 0 to 4). We have seen that $E(X) = 2.6$ and $E(Y) = 2.35$. These are *unconditional* expected values, for in computing these values we have not put any restrictions on them.

But now consider this question: What is the average number of printers sold (Y) if it is known that on a particular day 3 PCs were sold? Put differently, what is the *conditional expectation* of Y given that $X = 3$? Technically, what is $E(Y \mid X = 3)$? This is known as the conditional expectation of Y. Similarly, we could ask: What is $E(Y \mid X = 1)$?

From the preceding discussion it should be clear that in computing the unconditional expectation of an r.v., we do not take into account information about any other r.v., whereas in computing the conditional expectation we do.

To compute such conditional expectations, we use the following definition of conditional expectation

$$E(X \mid Y = y) = \sum_X X f(X \mid Y = y) \tag{B.34}$$

which gives the conditional expectation of X, where X is a discrete r.v., $f(X \mid Y = y)$ is the conditional PDF of X given in Eq. (A.20), and $\sum_X$ means the sum over all values of X. In relation to Equation (B.34), $E(X)$, considered earlier, is called the *unconditional expectation*. Computationally, $E(X \mid Y = y)$ is similar to $E(X)$ except that instead of using the unconditional PDF of X, we use its conditional PDF, as seen clearly in comparing Eq. (B.34) with Eq. (B.1).

Similarly,

$$E(Y \mid X = x) = \sum_Y Y f(Y \mid X = x) \tag{B.35}$$

gives the conditional expectation of Y. Let us illustrate with an example.

Example B.9.

Let us compute $E(Y \mid X = 2)$ for our PC/printer sales example. That is, we want to find out the conditional expected value of printers sold, knowing that 2 PCs have been sold per day. Using formula (B.34), we have

$$\begin{aligned} E(Y \mid X = 2) &= \sum_0^4 Y f(Y \mid x = 2) \\ &= f(Y = 1 \mid X = 2) + 2f(Y = 2 \mid X = 2) \\ &\quad + 3f(Y = 3 \mid X = 2) + 4f(Y = 4 \mid X = 2) \\ &= 1.875 \end{aligned}$$

Note: $f(Y = 1 \mid X = 2) = \frac{f(Y = 1, X = 2)}{f(X = 2)}$, and so on (see Table A-3).

As these calculations show, the conditional expectation of Y given that $X = 2$, is about 1.88, whereas, as shown previously, the unconditional expectation of Y was 2.35. Just as we saw previously that the conditional PDFs and marginal PDFs are generally different, the conditional and unconditional expectations in general are different too. Of course, if the two variables

are independent, the conditional and unconditional expectations will be the same. (Why?)

Conditional Variance

Just as we can compute the conditional expectation of a random variable, we can also compute its **conditional variance,** $\text{var}(Y \mid X)$. For Example B.9, for instance, we may be interested in finding the variance of Y, given that $X = 2$, $\text{var}(Y \mid X = 2)$. We can use formula (B.11) for the variance of X, except that we now have to use the conditional expectation of Y and the conditional PDF. To see how this is actually done, see Optional Exercise B.23. Incidentally, the variance formula given in Eq. (B.11) may be called the **unconditional variance** of X.

Just as conditional and unconditional expectations of an r.v., in general, are different, the conditional and unconditional variances, in general, are different also. They will be the same, however, if the two variables are independent.

As we will see in Chapter 2 and in subsequent chapters, the concepts of conditional expectation and conditional variance will play an important role in econometrics. Referring to the civilian labor force participation rate (CLFPR) and the civilian unemployment rate (CUNR) example discussed in Chapter 1, will the unconditional expectation of CLFPR be the same as the conditional expectation of CLFPR, conditioned on the knowledge of CUNR? If they are the same, then, the knowledge of CUNR is not particularly helpful in predicting CLFPR. In such a situation, regression analysis is not very useful. On the other hand, if the knowledge of CUNR enables us to forecast CLFPR better than without that knowledge, regression analysis becomes a very valuable research tool, as we show in the main chapters of the text.

B.6 SKEWNESS AND KURTOSIS

To conclude our discussion of the characteristics of probability distributions, we discuss the concepts of *skewness* and *kurtosis* of a probability distribution, which tell us something about the *shape* of the probability distribution. **Skewness (S)** is a measure of asymmetry, and **kurtosis (K)** is a measure of tallness or flatness of a PDF, as can be seen in Figure B-4.

To obtain measures of skewness and kurtosis, we need to know the *third moment* and the *fourth moment* of a PMF (PDF). We have already seen that the first moment of the PMF (PDF) of a random variable X is measured by $E(X) = \mu_X$, the mean of X, and the second moment around the mean (i.e., the variance) is measured by $E(X - \mu_x)^2$. In like fashion, the third and fourth moments around the mean value can be expressed as:

$$\text{Third moment: } E(X - \mu_x)^3 \tag{B.36}$$

$$\text{Fourth moment: } E(X - \mu_x)^4 \tag{B.37}$$

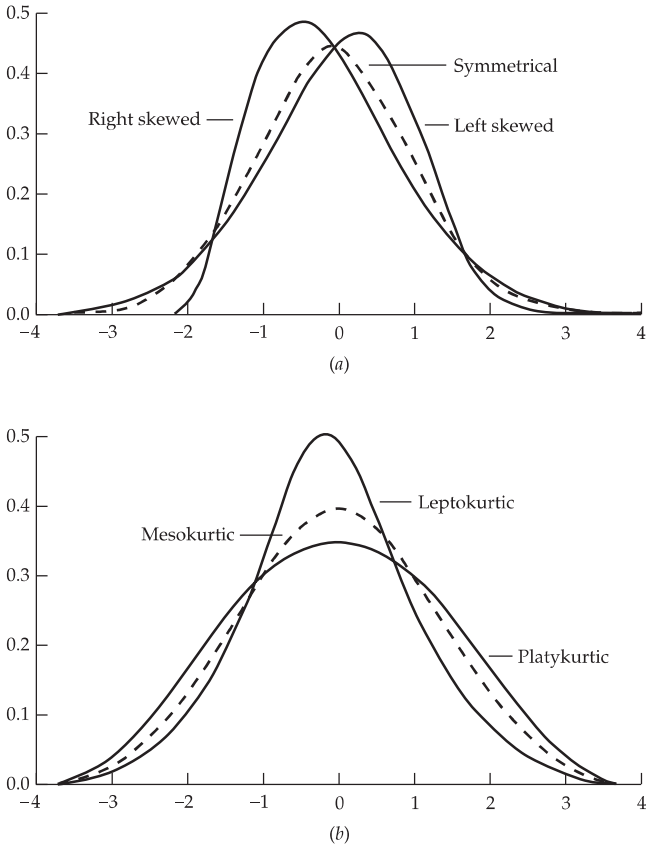

FIGURE B-4 (*a*) Skewness; (*b*) kurtosis

And, in general, the *r*th moment around the mean value can be expressed as

$$r\text{th moment: } E(X - \mu_x)^r \tag{B.38}$$

Given these definitions, the commonly used measures of skewness and kurtosis are as follows:

$$\begin{aligned} S &= \frac{E(X - \mu_x)^3}{\sigma_x^3} \\ &= \frac{\text{third moment about mean}}{\text{cube of standard deviation}} \end{aligned} \tag{B.39}$$

Since for a symmetrical PDF the third (and all odd order) moments are zero, for such a PDF the S value is zero. The prime example is the normal distribution, which we will discuss more fully in Appendix C. If the S value is positive, the

PDF is right-, or positively, skewed, and if it is negative, it is left-, or negatively, skewed. (See Fig. B-4[*a*]).

$$\text{K} = \frac{E(X - \mu_X)^4}{[E(X - \mu_X)^2]^2} = \frac{\text{fourth moment}}{\text{square of second moment}} \tag{B.40}$$

PDFs with values of K less than 3 are called *platykurtic* (fat or short-tailed), and those with values of K greater than 3 are called *leptokurtic* (slim or long-tailed), as shown in Fig. B-4(*b*). For a normal distribution the K value is 3, and such a PDF is called *mesokurtic*.

Since we will be making extensive use of the normal distribution in the main text, the knowledge that for such a distribution the values of S and K are zero and 3, respectively, will help us to compare other PDFs with the normal distribution.

The computational formulas to obtain the third and fourth moments of a PDF are straightforward extensions of the formula given in Eq. (B.11), namely,

$$\text{Third moment: } \sum (X - \mu_x)^3 f(X) \tag{B.41}$$

$$\text{Fourth moment: } \sum (X - \mu_x)^4 f(X) \tag{B.42}$$

where X is a discrete r.v. For a continuous r.v. we will replace the summation sign by the integral sign ($\int$).

Example B.10.

Consider the PDF given in Table B-1. For this PDF we have already seen that $E(X) = 3.5$ and Var $(X) = 2.9167$. The calculations of the third and fourth moments about the mean value are as follows:

X	$f(X)$	$(X-\mu_X)^3 f(X)$	$(X-\mu_X)^4 f(X)$
1	1/6	$(1-3.5)^3$ (1/6)	$(1-3.5)^4$ (1/6)
2	1/6	$(2-3.5)^3$ (1/6)	$(2-3.5)^4$ (1/6)
3	1/6	$(3-3.5)^3$ (1/6)	$(3-3.5)^4$ (1/6)
4	1/6	$(4-3.5)^3$ (1/6)	$(4-3.5)^4$ (1/6)
5	1/6	$(5-3.5)^3$ (1/6)	$(5-3.5)^4$ (1/6)
6	1/6	$(6-3.5)^3$ (1/6)	$(6-3.5)^4$ (1/6)
	Sum =	0	14.732

From the definitions of skewness and kurtosis given before, verify that for the present example the skewness coefficient is zero (Is that surprising?) and that the kurtosis value is 1.7317. Therefore, although the PDF given above is symmetrical around its mean value, it is *platykurtic*, or much flatter than the normal distribution, which should be apparent from its shape in Fig. B-4(*b*).

B.7 FROM THE POPULATION TO THE SAMPLE

To compute the characteristics of probability distributions, such as the expected value, variance, covariance, correlation coefficient, and conditional expected value, we obviously need the PMF (PDF), that is, the whole sample space or population. Thus, to find out the average income of all the people living in New York City at a given time, obviously we need information on the population of the whole city. Although conceptually there is some finite population of New York City at any given time, it is simply not practical to collect information about each member of the population (i.e., outcome, in the language of probability). What is done in practice is to draw a "representative" or a "random" sample from this population and to compute the average income of the people sampled.[5]

But will the average income obtained from the sample be equal to the true average income (i.e., expected value of income) in the population as a whole? Most likely it will not. Similarly, if we were to compute the variance of the income in the sampled population, would that equal the true variance that we would have obtained had we studied the whole population? Again, most likely it would not.

How then could we learn about population characteristics like the expected value, variance, etc., if we only have one or two samples from a given population? And, as we will see throughout the main chapters of the book, in practice, invariably we have to depend on one or more samples from a given population.

The answer to this very important question will be the focus of our attention in Appendix D. But meanwhile, we must find the sample counterparts, the **sample moments,** of the various population characteristics that we discussed in the preceding sections.

Sample Mean

Let X denote the number of cars sold per day by a car dealer. Assume that the r.v. X follows some PDF. Further, suppose we want to find out the average number [i.e., $E(X)$] of cars sold by the dealer in the first 10 days of each month. Assume that the car dealer has been in business for 10 years but has no time to look up the sales figures for the first 10 days of each month for the past 10 years. Suppose that he decides to pick at random the past data for one month and notes the sales figures for the first 10 days of that month, which are as follows: 9, 11, 11, 14, 13, 9, 8, 9, 14, and 12. This is a sample of 10 values. Notice that he has data for 120 months, and if he had decided to choose another month, he probably would have obtained 10 other values.

If the dealer adds up the 10 sales values and divides the sum by 10 (i.e., the sample size), the number he would obtain is known as the **sample mean.**

[5]The precise meaning of a random sample will be explained in Appendix C.

The sample mean of an r.v. X is generally denoted by the symbol $\overline{X}$ (read as X bar) and is defined as

$$\overline{X} = \sum_{i=1}^{n} \frac{X_i}{n} \tag{B.43}$$

where $\sum_{i=1}^{n} X_i$, as usual, means the sum of the X values from 1 to n, where n is the sample size.

The sample mean thus defined is known as an **estimator** of $E(X)$, which we can now call the *population mean.* An *estimator is simply a rule or formula that tells us how to go about estimating a population quantity,* such as the population mean. In Appendix D we will show how $\overline{X}$ is related to $E(X)$.

For the sample just given, the sample mean is

$$\overline{X} = \frac{9 + 11 + 11 + \cdots + 12}{10} = \frac{110}{10} = 11$$

which we call an **estimate** of the population mean. *An estimate is simply the numerical value taken by an estimator,* 11 in the preceding example. In our example, the average number of cars sold in the first 10 days of the month is 11. But keep in mind that this number will not necessarily equal $E(X)$; to compute the latter, we will have to take into account the sales data for the first 10 days of each of the other 119 months. In short, we will have to consider the entire PDF of car sales. But as we show in Appendix D, often the estimate, such as 11, obtained from a given sample is a fairly good "proxy" for the true $E(X)$.

Sample Variance

The ten sample values given previously are not all equal to the sample mean of 11. The variability of the ten values from this sample mean can be measured by the **sample variance,** denoted by S_x^2, which is an *estimator* of σ_x^2, *which we can now call the population variance.* The sample variance is defined as

$$S_x^2 = \sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{n - 1} \tag{B.44}$$

which is simply the sum of the squared difference of an individual X value from its (sample) mean value divided by the total number of observations less one.[6] The expression $(n - 1)$ is known as the **degrees of freedom,** whose precise meaning will be explained in Appendix C. S_x, the positive square root of S_x^2, is called the **sample standard deviation (sample s.d.).**

[6]If the sample size is reasonably large, we can divide by n instead of $(n - 1)$.

For the sample of 10 X values given earlier, the sample variance is

$$S_x^2 = \frac{(9-11)^2 + (11-11)^2 + \cdots + (12-11)^2}{9}$$

$$= \frac{44}{9} = 4.89$$

and the sample s.d. is $S_x = \sqrt{4.89} \approx 2.21$. Note that 4.89 is an *estimate* of the population variance and 2.21 is an *estimate* of the population s.d. Again, an estimate is a numerical value taken by an estimator in a given sample.

Sample Covariance

Example B.11.

Suppose we have a bivariate population of two variables Y (stock prices) and X (consumer prices). Suppose further that from this bivariate population we obtain the random sample shown in the first two columns of Table B-3. In this example, stock prices are measured by the Dow Jones average and consumer prices by the Consumer Price Index (CPI). The other entries in this table are discussed later.

TABLE B-3 SAMPLE COVARIANCE AND SAMPLE CORRELATION COEFFICIENT BETWEEN DOW JONES AVERAGE (Y) AND CONSUMER PRICE INDEX (X) OVER THE PERIOD 1998–2007

Year	*Dow* Y (1)	*CPI* X (2)	$(Y-\bar{Y})(X-\bar{X})$ (3)
1998	8,625.52	163.00	(8625.5 − 10367.8)(163 − 183.6)
1999	10,464.88	166.60	(10464.9 − 10367.8)(166.6 − 183.6)
2000	10,734.90	172.20	—
2001	10,189.13	177.10	—
2002	9,226.43	179.90	—
2003	8,993.59	184.00	—
2004	10,317.39	188.90	—
2005	10,547.67	195.30	—
2006	11,408.67	201.60	(11408.7 − 10367.8)(201.6 − 183.6)
2007	13,169.98	207.34	(13170 − 10367.8)(207.3 − 183.6)
Sum	103,678.16	1,835.94	≈121,992.73

$\bar{Y} = \frac{103{,}678.16}{10} = 10367.8$ Sample var(Y) = 1,708,150

$\bar{X} = \frac{1{,}835.94}{10} = 183.594$ Sample var(X) = 216.898

Source: Data on X and Y are from the *Economic Report of the President,* 2008, Tables B-95, B-96, and B-60, respectively.

Analogous to the population covariance defined in Eq. (B.23), the **sample covariance** between two random variables X and Y is defined as

$$\text{Sample cov } (X, Y) = \frac{\Sigma(X_i - \overline{X})(Y_i - \overline{Y})}{n - 1} \tag{B.45}$$

which is simply the sum of the cross products of the two random variables expressed as deviations from their (sample) mean values and divided by the degrees of freedom, $(n - 1)$. (If the sample size is large, we may use n as the divisor.) The sample covariance defined in Equation (B.45) is thus the estimator of the population covariance. Its numerical value in a given instance will provide an *estimate* of the population covariance, as in the following example.

In Table B-3 we have given the necessary quantities to compute the sample covariance, which in the present case is

$$\text{Sample cov } (X, Y) = \frac{121{,}992.73}{9} = 13{,}554.75$$

Thus, in the present case the covariance between stock prices and consumer prices is positive. Some analysts believe that investment in stocks is a hedge against inflation; that is, as inflation increases, stock prices increase, too. Apparently, for the period 1998 to 2007 that seems to be the case, although empirical evidence on this subject is not unequivocal.

Sample Correlation Coefficient

In Eq. (B.29) we defined the population correlation coefficient between two random variables. Its sample analogue, or estimator, which we denote by the symbol r, is as follows:

$$\begin{aligned} r &= \frac{\Sigma_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})/(n - 1)}{S_x S_y} \\ &= \frac{\text{sample cov } (X, Y)}{\text{s.d.}(X)\text{s.d.}(Y)} \end{aligned} \tag{B.46}$$

The **sample correlation** coefficient thus defined has the same properties as the population correlation coefficient ρ; they both lie between -1 and $+1$.

For the data given in Table B-3 you can easily compute the sample standard deviations of Y and X, and therefore can compute the sample correlation coefficient r, an estimate of ρ, which turns out to be

$$\begin{aligned} r &= \frac{13{,}554.75}{(14.727)(1306.962)} \\ &= 0.7042 \end{aligned}$$

Thus, in our example stock prices and consumer prices are pretty positively correlated because the computed value is close to 1.

Sample Skewness and Kurtosis

To compute **sample skewness** and **sample kurtosis** values, we use the sample third and fourth moments (compare with Eqs. [B.36] and [B.37]). The sample third moment (compare with the formula for sample variance) is

$$\frac{\sum(X - \overline{X})^3}{(n - 1)} \tag{B.47}$$

and the sample fourth moment is

$$\frac{\sum(X - \overline{X})^4}{(n - 1)} \tag{B.48}$$

Using the data given in Table B-3, calculate the sample third and fourth moments and divide them by the standard deviation value to the third and fourth powers, respectively. Verify that the sample skewness and kurtosis measures for the Dow Jones average are 0.6873 and 2.9447, respectively, suggesting that the distribution of the Dow Jones average is positively skewed and that it is flatter than a normal distribution.

B.8 SUMMARY

After introducing several fundamental concepts of probability, random variables, probability distributions, etc., in Appendix A, in this appendix we discussed some major characteristics or moments of probability distributions of random variables, such as the expected value, variance, covariance, correlation, skewness, kurtosis, conditional expectation, and conditional variance. We also discussed the famous Chebyshev's inequality. The discussion of these concepts has been somewhat intuitive, for our objective here is not to teach statistics per se but simply to review some of its major concepts that are needed to follow the various topics discussed in the main chapters of this book.

In this appendix we also presented several important formulas. These formulas tell us how to compute the probabilities of random variables and how to estimate the characteristics of probability distributions (i.e., the moments), such as the expected or mean value, variance, covariance, correlation, and conditional expectation. In presenting these formulas, we made a careful distinction between the *population moments* and *sample moments* and gave the appropriate computing formulas. Thus, $E(X)$, the expected value of the r.v. X, is a population moment, that is, the mean value of X if the entire population of the X values were known. On the other hand, $\overline{X}$ is a sample moment, that is, the average value of X if it is based on sample values of X and not on the entire population. In statistics the dichotomy between the population and the sample is very important, for in most applications we have only one or two samples from some population of interest and often we want to draw inferences about the population moments on the basis of the sample moments. We will explain how this is done in Appendixes C and D.

KEY TERMS AND CONCEPTS

The key terms and concepts introduced in this appendix are

Characteristics (moments) of univariate PMFs
- **a)** expected value (population mean value)
- **b)** variance
- **c)** standard deviation (s.d.)
- **d)** coefficient of variation (V)

Characteristics of multivariate PDFs
- **a)** covariance
- **b)** (population) coefficient of correlation
- **c)** correlation
- **d)** conditional expectation
- **e)** unconditional expectation
- **f)** conditional variance
- **g)** unconditional variance
- **h)** skewness (S)
- **i)** kurtosis (K)

Population vs. sample
- **a)** sample moments
- **b)** sample mean
- **c)** estimator; estimate
- **d)** sample variance
- **e)** degrees of freedom
- **f)** sample standard deviation (sample s.d.)
- **g)** sample covariance
- **h)** sample correlation
- **i)** sample skewness
- **j)** sample kurtosis

QUESTIONS

B.1. What is meant by the moments of a PDF? What are the most frequently used moments?

B.2. Explain the meaning of
- **a.** expected value
- **b.** variance
- **c.** standard deviation
- **d.** covariance
- **e.** correlation
- **f.** conditional expectation

B.3. Explain the meaning of
- **a.** sample mean
- **b.** sample variance
- **c.** sample standard deviation
- **d.** sample covariance
- **e.** sample correlation

B.4. Why is it important to make the distinction between population moments and sample moments?

B.5. Fill in the gaps in the manner of (*a*) below.
- **a.** The expected value or mean is a measure of central tendency.
- **b.** The variance is a measure of . . .
- **c.** The covariance is a measure of . . .
- **d.** The correlation is a measure of . . .

B.6. A random variable (r.v.) X has a mean value of \$50 and its standard deviation (s.d.) is \$5. Is it correct to say that its variance is \$25 squared? Why or why not?

B.7. Explain whether the following statements are true or false. Give reasons.
- **a.** Although the expected value of an r.v. can be positive or negative, its variance is always positive.

b. The coefficient of correlation will have the same sign as that of the covariance between the two variables.
c. The conditional and unconditional expectations of an r.v. mean the same thing.
d. If two variables are independent, their correlation coefficient will always be zero.
e. If the correlation coefficient between two variables is zero, it means that the two variables are independent.
f. $E\left(\frac{1}{X}\right) = \frac{1}{E(X)}$
g. $E[X - \mu_X]^2 = [E(X - \mu_X)]^2$

PROBLEMS

B.8. Refer to Problem A.12.
a. Find the expected value of X.
b. What is the variance and standard deviation of X?
c. What is the coefficient of variation of X?
d. Find the skewness and kurtosis values of X.

B.9. The following table gives the anticipated 1-year rates of return from a certain investment and their probabilities.

TABLE B-4 ANTICIPATED 1-YEAR RATE OF RETURN FROM A CERTAIN INVESTMENT

Rate of return (X) %	$f(X)$
−20	0.10
−10	0.15
10	0.45
25	0.25
30	0.05
Total	1.00

a. What is the expected rate of return from this investment?
b. Find the variance and standard deviation of the rate of return.
c. Find the skewness and kurtosis coefficients.
d. Find the cumulative distribution function (CDF) and obtain the probability that the rate of return is 10 percent or less.

B.10. The following table gives the joint PDF of random variables X and Y, where X = the first-year rate of return (%) expected from investment A, and Y = the first-year rate of return (%) expected from investment B.

TABLE B-5 RATES OF RETURN ON TWO INVESTMENTS

	X (%)			
Y(%)	−10	0	20	30
20	0.27	0.08	0.16	0.00
50	0.00	0.04	0.10	0.35

a. Find the marginal distributions of Y and X.
b. Calculate the expected rate of return from investment B.
c. Find the conditional distribution of Y, given $X = 20$.
d. Are X and Y independent random variables? How do you know? Hint:

$$E(XY) = \sum_{X=1}^{4} \sum_{Y=1}^{2} X_i Y_i f(X_i, Y_i)$$

B.11. You are told that $E(X) = 8$ and var $(X) = 4$. What are the expected values and variances of the following expressions?
a. $Y = 3X + 2$
b. $Y = 0.6X - 4$
c. $Y = X/4$
d. $Y = aX + b$, where a and b are constants
e. $Y = 3X^2 + 2$
How would you express these formulas verbally?

B.12. Consider formulas (B.32) and (B.33). Let X stand for the rate of return on a security, say, IBM, and Y the rate of return on another security, say, General Foods. Let $s_X^2 = 16$, $s_Y^2 = 9$, and $r = -0.8$. What is the variance of $(X + Y)$ in this case? Is it greater than or smaller than var (X) + var (Y)? In this instance, is it better to invest equally in the two securities (i.e., diversify) than in either security exclusively? This problem is the essence of the *portfolio theory* of finance. (See, for example, Richard Brealey and Stewart Myers, *Principles of Corporate Finance,* McGraw-Hill, New York, latest edition.)

B.13. Table B-6 gives data on the number of new business incorporations (Y) and the number of business failures (X) for the United States from 1984 to 1995.
a. What is the average value of new business incorporations? And the variance?
b. What is the average value of business failures? And the variance?
c. What is the covariance between Y and X? And the correlation coefficient?

TABLE B-6 NUMBER OF NEW BUSINESS INCORPORATIONS (Y) AND NUMBER OF BUSINESS FAILURES (X), UNITED STATES, 1984–1995

YEAR	Y	X
1984	634,991	52,078
1985	664,235	57,253
1986	702,738	61,616
1987	685,572	61,111
1988	685,095	57,097
1989	676,565	50,361
1990	647,366	60,747
1991	628,604	88,140
1992	666,800	97,069
1993	706,537	86,133
1994	741,778	71,558
1995	766,988	71,128

Source: Economic Report of the President, 2004, Table B-96, p. 395.

d. Are the two variables independent?
e. If there is correlation between the two variables, does this mean that one variable causes the other variable? That is, do new incorporations cause business failures, or vice versa?

B.14. For Problem A.13, find out the var $(X + Y)$. How would you interpret this variance?

B.15. Refer to Table 1-2 given in Problem 1.6.
a. Compute the covariances between the S&P 500 index and the CPI and between the three-month Treasury bill rate and the CPI. Are these population or sample covariances?
b. Compute the correlation coefficients between the S&P 500 index and the CPI and between the three-month Treasury bill rate and the CPI. A priori, would you expect these correlation coefficients to be positive or negative? Why?
c. If there is a positive relationship between the CPI and the three-month Treasury bill rate, does that mean *inflation*, as measured by the CPI, is the *cause* of higher T bill rates?

B.16. Refer to Table 1-3 in Problem 1.7. Let ER stand for U.K. pound/$ exchange rate (i.e., the number of U.K. pounds per U.S. dollar) and RPR stand for the ratio of the U.S. CPI/U.K. CPI. Is the correlation between ER and RPR expected to be positive or negative? Why? Show your computations. Would your answer change if you found correlation between ER and (1/RPR)? Why?

OPTIONAL EXERCISES

B.17. Find the expected value of the following PDF:

$$f(X) = \frac{X^2}{9} \qquad 0 \leq x \leq 3$$

B.18. Show that
a. $E(X^2) \geq [E(X)]^2$ *Hint:* Recall the definition of variance.
b. $\text{cov}\,(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$
$= E(XY) - \mu_X\mu_Y$
where $\mu_X = E(X)$ and $\mu_Y = E(Y)$.
How would you express these formulas verbally?

B.19. Establish Eq. (B.15). *Hint:* $\text{Var}\,(aX) = \text{E}[aX - \text{E}(aX)]^2$ and simplify.

B.20. Establish Eq. (B.17). *Hint:* $\text{Var}\,(aX + bY) = \text{E}[(aX + bY) - \text{E}(aX + bY)]^2$ and simplify.

B.21. According to Chebyshev's inequality, what percentage of any set of data must lie within c standard deviations on either side of the mean value if (a) $c = 2.5$ and (b) $c = 8$?

B.22. Show that $E(X - k)^2 = \text{var}\,(X) + [E(X) - k]^2$. For what value of k will $E(X - k)^2$ be minimum? And what is that value of k?

B.23. For the PC/printer sales example discussed in this appendix compute the conditional variance of Y (printers sold) given that X (PCs sold) is 2. *Hint:* Use the conditional expectation given in Example B.9 and use the formula:

$$\text{var}\,(Y \mid X = 2) = \sum [Y_i - E(Y \mid X = 2)]^2 f(Y \mid X = 2)$$

B.24. Compute the expected value and variance for the PDF given in Problem A.19.