CHAPTER 4

Review of Basic Probability and Statistics

Recommended sections for a first reading: 4.1 through 4.7

4.1 INTRODUCTION

The completion of a successful simulation study involves much more than constructing a flowchart of the system under study, translating the flowchart into a computer "program," and then making one or a few replications of each proposed system configuration. The use of probability and statistics is such an integral part of a simulation study that every simulation modeling team should include at least one person who is thoroughly trained in such techniques. In particular, probability and statistics are needed to understand how to model a probabilistic system (see Sec. 4.7), validate the simulation model (Chap. 5), choose the input probability distributions (Chap. 6), generate random samples from these distributions (Chaps. 7 and 8), perform statistical analyses of the simulation output data (Chaps. 9 and 10), and design the simulation experiments (Chaps. 11 and 12).

In this chapter we establish statistical notation used throughout the book and review some basic probability and statistics particularly relevant to simulation. We also point out the potential dangers of applying classical statistical techniques based on independent observations to simulation output data, which are rarely, if ever, independent.

4.2 RANDOM VARIABLES AND THEIR PROPERTIES

An *experiment* is a process whose outcome is not known with certainty. The set of all possible outcomes of an experiment is called the *sample space* and is denoted by S. The outcomes themselves are called the *sample points* in the sample space.

EXAMPLE 4.1. If the experiment consists of flipping a coin, then

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

where the symbol { } means the "set consisting of," and "H" and "T" mean that the outcome is a head and a tail, respectively.

EXAMPLE 4.2. If the experiment consists of tossing a die, then

$$S = \{1, 2, \ldots, 6\}$$

where the outcome i means that i appeared on the die, $i = 1, 2, \ldots, 6$.

A *random variable* is a function (or rule) that assigns a real number (any number greater than $-\infty$ and less than ∞) to each point in the sample space S.

EXAMPLE 4.3. Consider the experiment of rolling a pair of dice. Then

$$S = \{(1, 1), (1, 2), \ldots, (6, 6)\}$$

where (i, j) means that i and j appeared on the first and second die, respectively. If X is the random variable corresponding to the sum of the two dice, then X assigns the value 7 to the outcome (4, 3).

EXAMPLE 4.4. Consider the experiment of flipping two coins. If X is the random variable corresponding to the number of heads that occur, then X assigns the value 1 to either the outcome (H, T) or the outcome (T, H).

In general, we denote random variables by capital letters such as X, Y, Z and the values that random variables take on by lowercase letters such as x, y, z.

The *distribution function* (sometimes called the *cumulative* distribution function) $F(x)$ of the random variable X is defined for each real number x as follows:

$$F(x) = P(X \leq x) \qquad \text{for } -\infty < x < \infty$$

where $P(X \leq x)$ means the probability associated with the event $\{X \leq x\}$. [See Ross (1997, chap. 1) for a discussion of events and probabilities.] Thus, $F(x)$ is the probability that, when the experiment is done, the random variable X will have taken on a value no larger than the number x.

A distribution function $F(x)$ has the following properties:

1. $0 \leq F(x) \leq 1$ for all x.
2. $F(x)$ is nondecreasing [i.e., if $x_1 < x_2$, then $F(x_1) \leq F(x_2)$].
3. $\lim_{x \to \infty} F(x) = 1$ and $\lim_{x \to -\infty} F(x) = 0$ (since X takes on only finite values).

A random variable X is said to be *discrete* if it can take on at most a countable number of values, say, $x_1, x_2, \ldots$. ("Countable" means that the set of possible values can be put in a one-to-one correspondence with the set of positive integers. An example of an uncountable set is all real numbers between 0 and 1.) Thus, a random variable that takes on only a finite number of values $x_1, x_2, \ldots, x_n$ is discrete. The probability that the discrete random variable X takes on the value x_i is given by

$$p(x_i) = P(X = x_i) \qquad \text{for } i = 1, 2, \ldots$$

and we must have

$$\sum_{i=1}^{\infty} p(x_i) = 1$$

where the summation means add together $p(x_1)$, $p(x_2)$, ... All probability statements about X can be computed (at least in principle) from $p(x)$, which is called the *probability mass function* for the discrete random variable X. If $I = [a, b]$, where a and b are real numbers such that $a \leq b$, then

$$P(X \in I) = \sum_{a \leq x_i \leq b} p(x_i)$$

where the symbol $\in$ means "contained in" and the summation means add together $p(x_i)$ for all x_i such that $a \leq x_i \leq b$. The distribution function $F(x)$ for the discrete random variable X is given by

$$F(x) = \sum_{x_i \leq x} p(x_i) \qquad \text{for all } -\infty < x < \infty$$

EXAMPLE 4.5. For the inventory example of Sec. 1.5, the size of the demand for the product is a discrete random variable X that takes on the values 1, 2, 3, 4 with respective probabilities $\frac{1}{6}, \frac{1}{3}, \frac{1}{3}, \frac{1}{6}$. The probability mass function and the distribution function for X are given in Figs. 4.1 and 4.2. Furthermore,

$$P(2 \leq X \leq 3) = p(2) + p(3) = \tfrac{1}{3} + \tfrac{1}{3} = \tfrac{2}{3}$$

EXAMPLE 4.6. A manufacturing system produces parts that then must be inspected for quality. Suppose that 90 percent of the inspected parts are good (denoted by 1) and 10 percent are bad and must be scrapped (denoted by 0). If X denotes the outcome of inspecting a part, then X is a discrete random variable with $p(0) = 0.1$ and $p(1) = 0.9$. (See the discussion of the Bernoulli random variable in Sec. 6.2.3.)

We now consider random variables that can take on an uncountably infinite number of different values (e.g., all nonnegative real numbers). A random variable X is said to be *continuous* if there exists a nonnegative function $f(x)$ such that for

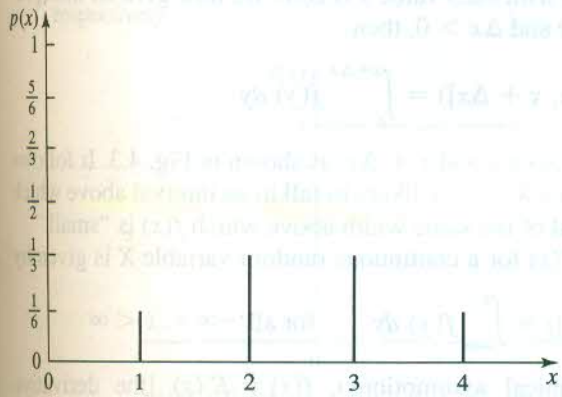

FIGURE 4.1
$p(x)$ for the demand-size random variable X.

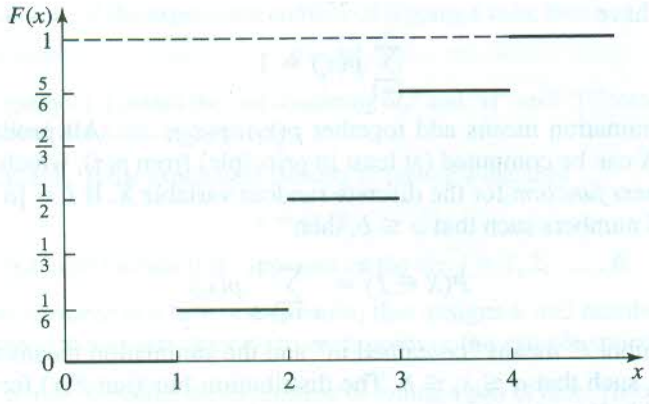

FIGURE 4.2
$F(x)$ for the demand-size random variable X.

any set of real numbers B (e.g., B could be all real numbers between 1 and 2),

$$P(X \in B) = \int_B f(x)\,dx \qquad \text{and} \qquad \int_{-\infty}^{\infty} f(x)\,dx = 1$$

[Thus, the total area under $f(x)$ is 1. Also, if X is a nonnegative random variable as is often the case in simulation applications, the second range of integration is from 0 to ∞.] All probability statements about X can (in principle) be computed from $f(x)$, which is called the *probability density function* for the continuous random variable X.

For a discrete random variable X, $p(x)$ is the actual probability associated with the value x. However, $f(x)$ is *not* the probability that a continuous random variable X equals x. For any real number x,

$$P(X = x) = P(X \in [x, x]) = \int_x^x f(y)\,dy = 0$$

Since the probability associated with each value x is zero, we now give an interpretation to $f(x)$. If x is any number and $\Delta x > 0$, then

$$P(X \in [x, x + \Delta x]) = \int_x^{x+\Delta x} f(y)\,dy$$

which is the area under $f(x)$ between x and $x + \Delta x$, as shown in Fig. 4.3. It follows that a continuous random variable X is more likely to fall in an interval above which $f(x)$ is "large" than in an interval of the same width above which $f(x)$ is "small."

The distribution function $F(x)$ for a continuous random variable X is given by

$$F(x) = P(X \in (-\infty, x]) = \int_{-\infty}^{x} f(y)\,dy \qquad \text{for all } -\infty < x < \infty$$

Thus (under some mild technical assumptions), $f(x) = F'(x)$ [the derivative of $F(x)$]. Furthermore, if $I = [a, b]$ for any real numbers a and b such that $a < b$,

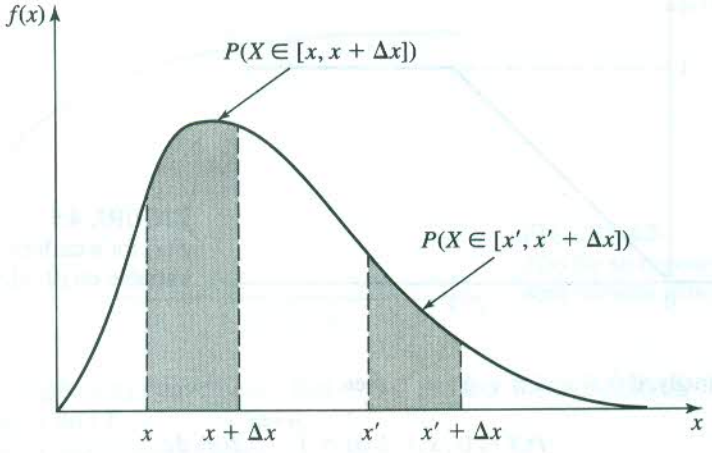

FIGURE 4.3
Interpretation of the probability density function $f(x)$.

then

$$P(X \in I) = \int_a^b f(y)\,dy = F(b) - F(a)$$

where the last equality is an application of the *fundamental theorem of calculus*, since $F'(x) = f(x)$.

EXAMPLE 4.7. A uniform random variable on the interval [0, 1] has the following probability density function:

$$f(x) = \begin{cases} 1 & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Furthermore, if $0 \le x \le 1$, then

$$F(x) = \int_0^x f(y)\,dy = \int_0^x 1\,dy = x$$

[What is $F(x)$ if $x < 0$ or if $x > 1$?] Plots of $f(x)$ and $F(x)$ are given in Figs. 4.4 and 4.5, respectively.

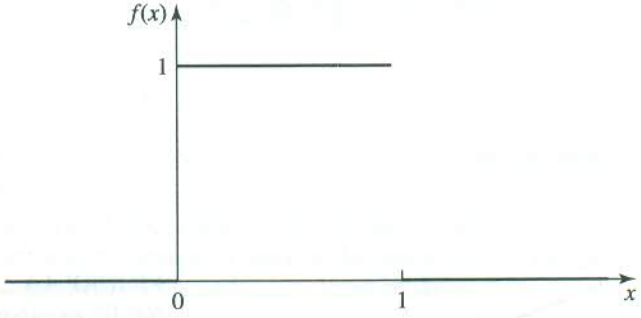

FIGURE 4.4
$f(x)$ for a uniform random variable on [0, 1].

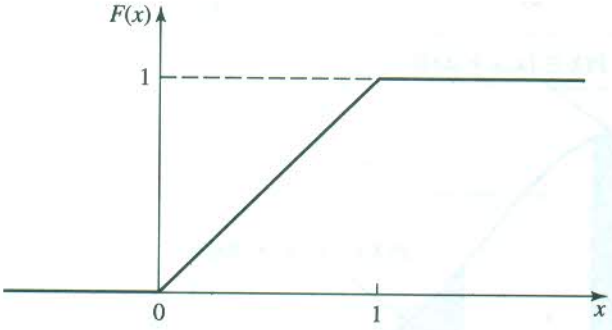

FIGURE 4.5
$F(x)$ for a uniform random variable on [0, 1].

Finally, if $0 \le x < x + \Delta x \le 1$, then

$$\begin{aligned} P(X \in [x, x + \Delta x]) &= \int_x^{x+\Delta x} f(y)\, dy \\ &= F(x + \Delta x) - F(x) \\ &= (x + \Delta x) - x \\ &= \Delta x \end{aligned}$$

It follows that a uniform random variable on [0, 1] is equally likely to fall in any interval of length Δx between 0 and 1, which justifies the name "uniform." The uniform random variable on [0, 1] is fundamental to simulation, since it is the basis for generating any random quantity on a computer (see Chaps. 7 and 8).

EXAMPLE 4.8. In Chap. 1 the exponential random variable was used for interarrival and service times in the queueing example and for interdemand times in the inventory example. The probability density function and distribution function for an exponential random variable with mean β are given in Figs. 4.6 and 4.7.

So far in this chapter we have considered only one random variable at a time, but in a simulation one must usually deal with n (a positive integer) random variables $X_1, X_2, \ldots, X_n$ simultaneously. For example, in the queueing model of Sec. 1.4, we were interested in the (input) service-time random variables $S_1, S_2, \ldots, S_n$ and the (output) delay random variables $D_1, D_2, \ldots, D_n$. In the discussion that follows, we

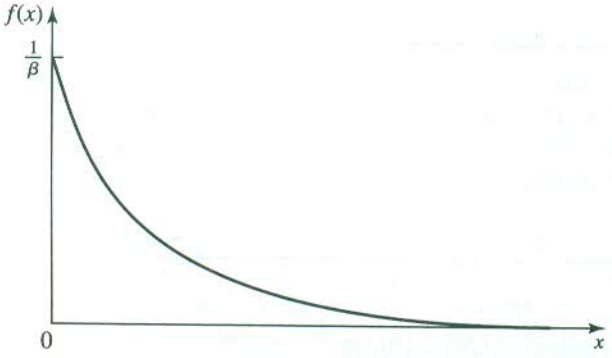

FIGURE 4.6
$f(x)$ for an exponential random variable with mean β.

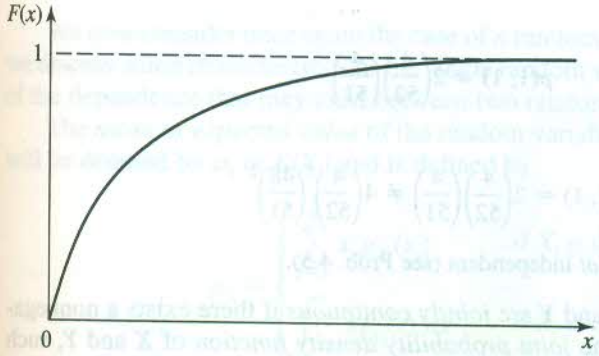

FIGURE 4.7
$F(x)$ for an exponential random variable with mean β.

assume for expository convenience that $n = 2$ and that the two random variables in question are X and Y.

If X and Y are discrete random variables, then let

$$p(x, y) = P(X = x, Y = y) \qquad \text{for all } x, y$$

where $p(x, y)$ is called the *joint probability mass function* of X and Y. In this case, X and Y are *independent* if

$$p(x, y) = p_X(x)p_Y(y) \qquad \text{for all } x, y$$

where

$$p_X(x) = \sum_{\text{all } y} p(x, y)$$

$$p_Y(y) = \sum_{\text{all } x} p(x, y)$$

are the (marginal) probability mass functions of X and Y.

EXAMPLE 4.9. Suppose that X and Y are jointly discrete random variables with

$$p(x, y) = \begin{cases} \dfrac{xy}{27} & \text{for } x = 1, 2 \text{ and } y = 2, 3, 4 \\ 0 & \text{otherwise} \end{cases}$$

Then

$$p_X(x) = \sum_{y=2}^{4} \frac{xy}{27} = \frac{x}{3} \qquad \text{for } x = 1, 2$$

$$p_Y(y) = \sum_{x=1}^{2} \frac{xy}{27} = \frac{y}{9} \qquad \text{for } y = 2, 3, 4$$

Since $p(x, y) = xy/27 = p_X(x)p_Y(y)$ for all x, y, the random variables X and Y are independent.

EXAMPLE 4.10. Suppose that 2 cards are dealt from a deck of 52 without replacement. Let the random variables X and Y be the number of aces and kings that occur, both of which have possible values of 0, 1, 2. It can be shown that

$$p_X(1) = p_Y(1) = 2\left(\frac{4}{52}\right)\left(\frac{48}{51}\right)$$

and

$$p(1, 1) = 2\left(\frac{4}{52}\right)\left(\frac{4}{51}\right)$$

Since

$$p(1, 1) = 2\left(\frac{4}{52}\right)\left(\frac{4}{51}\right) \neq 4\left(\frac{4}{52}\right)^2\left(\frac{48}{51}\right)^2$$

it follows that X and Y are *not* independent (see Prob. 4.5).

The random variables X and Y are *jointly continuous* if there exists a nonnegative function $f(x, y)$, called the *joint probability density function* of X and Y, such that for all sets of real numbers A and B,

$$P(X \in A, Y \in B) = \int_B \int_A f(x, y)\, dx\, dy$$

In this case, X and Y are *independent* if

$$f(x, y) = f_X(x) f_Y(y) \qquad \text{for all } x, y$$

where

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx$$

are the (marginal) probability density functions of X and Y, respectively.

EXAMPLE 4.11. Suppose that X and Y are jointly continuous random variables with

$$f(x, y) = \begin{cases} 24xy & \text{for } x \geq 0,\ y \geq 0, \text{ and } x + y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Then

$$f_X(x) = \int_0^{1-x} 24xy\, dy = 12xy^2 \Big|_0^{1-x} = 12x(1 - x)^2 \qquad \text{for } 0 \leq x \leq 1$$

$$f_Y(y) = \int_0^{1-y} 24xy\, dx = 12yx^2 \Big|_0^{1-y} = 12y(1 - y)^2 \qquad \text{for } 0 \leq y \leq 1$$

Since

$$f\left(\frac{1}{2}, \frac{1}{2}\right) = 6 \neq \left(\frac{3}{2}\right)^2 = f_X\left(\frac{1}{2}\right) f_Y\left(\frac{1}{2}\right)$$

X and Y are not independent.

Intuitively, the random variables X and Y (whether discrete or continuous) are independent if knowing the value that one random variable takes on tells us nothing about the distribution of the other. Also, if X and Y are not independent, we say that they are *dependent*.

We now consider once again the case of n random variables $X_1, X_2, \ldots, X_n$, and we discuss some characteristics of the single random variable X_i and some measures of the dependence that may exist between two random variables X_i and X_j.

The *mean* or *expected value* of the random variable X_i (where $i = 1, 2, \ldots, n$) will be denoted by μ_i or $E(X_i)$ and is defined by

$$\mu_i = \begin{cases} \sum_{j=1}^{\infty} x_j p_{X_i}(x_j) & \text{if } X_i \text{ is discrete} \\ \int_{-\infty}^{\infty} x f_{X_i}(x)\, dx & \text{if } X_i \text{ is continuous} \end{cases}$$

The mean is one measure of central tendency in the sense that it is the center of gravity [see, for example, Billingsley et al. (1986, pp. 42–43)].

EXAMPLE 4.12. For the demand-size random variable in Example 4.5, the mean is given by

$$\mu = 1\left(\frac{1}{6}\right) + 2\left(\frac{1}{3}\right) + 3\left(\frac{1}{3}\right) + 4\left(\frac{1}{6}\right) = \frac{5}{2}$$

EXAMPLE 4.13. For the uniform random variable in Example 4.7, the mean is given by

$$\mu = \int_0^1 x f(x)\, dx = \int_0^1 x\, dx = \frac{1}{2}$$

Let c or c_i denote a constant (real number). Then the following are important properties of means:

1. $E(cX) = cE(X)$.
2. $E(\Sigma_{i=1}^n c_i X_i) = \Sigma_{i=1}^n c_i E(X_i)$ *even if the* X_i*'s are dependent.*

The *median* $x_{0.5}$ of the random variable X_i, which is an alternative measure of central tendency, is defined to be the smallest value of x such that $F_{X_i}(x) \geq 0.5$. If X_i is a continuous random variable, then $F_{X_i}(x_{0.5}) = 0.5$, as shown in Fig. 4.8. The median may be a better measure of central tendency than the mean when X_i can take on very large or very small values, since extreme values can greatly affect the mean even if they are very unlikely to occur; such is not the case with the median.

EXAMPLE 4.14. Consider a discrete random variable X that takes on each of the values, 1, 2, 3, 4, and 5 with probability 0.2. Clearly, the mean and median of X are 3. Consider now a random variable Y that takes on each of the values 1, 2, 3, 4, and 100 with probability 0.2. The mean and median of Y are 22 and 3, respectively. Note that the median is insensitive to this change in the distribution.

The *variance* of the random variable X_i will be denoted by σ_i^2 or $\text{Var}(X_i)$ and is defined by

$$\sigma_i^2 = E[(X_i - \mu_i)^2] = E(X_i^2) - \mu_i^2$$

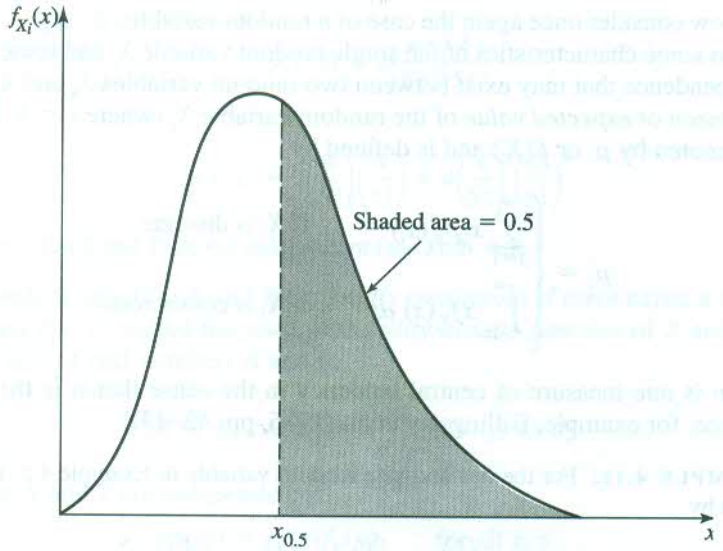

FIGURE 4.8
The median $x_{0.5}$ for a continuous random variable.

The variance is a measure of the dispersion of a random variable about its mean, as seen in Fig. 4.9. The larger the variance, the more likely the random variable is to take on values far from its mean.

EXAMPLE 4.15. For the demand-size random variable in Example 4.5, the variance is computed as follows:

$$E(X^2) = 1^2\left(\frac{1}{6}\right) + 2^2\left(\frac{1}{3}\right) + 3^2\left(\frac{1}{3}\right) + 4^2\left(\frac{1}{6}\right) = \frac{43}{6}$$

$$\text{Var}(X) = E(X^2) - \mu^2 = \frac{43}{6} - \left(\frac{5}{2}\right)^2 = \frac{11}{12}$$

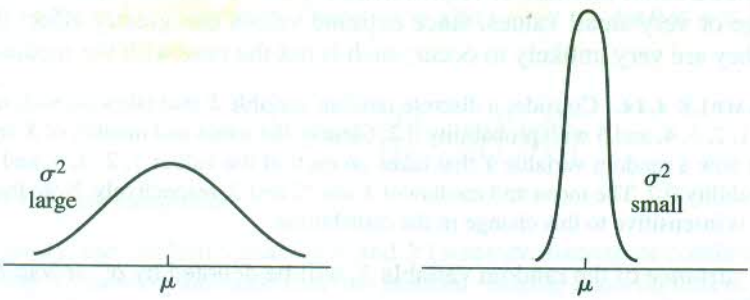

FIGURE 4.9
Density functions for continuous random variables with large and small variances.

EXAMPLE 4.16. For the uniform random variable on [0, 1] in Example 4.7, the variance is computed as follows:

$$E(X^2) = \int_0^1 x^2 f(x)\,dx = \int_0^1 x^2\,dx = \frac{1}{3}$$

$$\text{Var}(X) = E(X^2) - \mu^2 = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12}$$

The variance has the following properties:

1. $\text{Var}(X) \geq 0$.
2. $\text{Var}(cX) = c^2\,\text{Var}(X)$.
3. $\text{Var}(\Sigma_{i=1}^n X_i) = \Sigma_{i=1}^n \text{Var}(X_i)$ if the X_i's are independent (or uncorrelated, as discussed below).

The *standard deviation* of the random variable X_i is defined to be $\sigma_i = \sqrt{\sigma_i^2}$. The standard deviation can be given the most definitive interpretation when X_i has a normal distribution (see Sec. 6.2.2). In particular, suppose that X_i has a normal distribution with mean μ_i and standard deviation σ_i. In this case, for example, the probability that X_i is between $\mu_i - 1.96\sigma_i$ and $\mu_i + 1.96\sigma_i$ is 0.95.

We now consider measures of dependence between two random variables. The *covariance* between the random variables X_i and X_j (where $i = 1, 2, \ldots, n; j = 1, 2, \ldots, n$), which is a measure of their (linear) dependence, will be denoted by C_{ij} or $\text{Cov}(X_i, X_j)$ and is defined by

$$C_{ij} = E[(X_i - \mu_i)(X_j - \mu_j)] = E(X_i X_j) - \mu_i \mu_j \tag{4.1}$$

Note that covariances are symmetric, that is, $C_{ij} = C_{ji}$, and that if $i = j$, then $C_{ij} = C_{ii} = \sigma_i^2$.

EXAMPLE 4.17. For the jointly continuous random variables X and Y in Example 4.11, the covariance is computed as

$$E(XY) = \int_0^1 \int_0^{1-x} xyf(x, y)\,dy\,dx$$

$$= \int_0^1 x^2 \left(\int_0^{1-x} 24y^2\,dy \right) dx$$

$$= \int_0^1 8x^2(1 - x)^3\,dx$$

$$= \frac{2}{15}$$

$$E(X) = \int_0^1 x f_X(x)\,dx = \int_0^1 12x^2(1 - x)^2\,dx = \frac{2}{5}$$

$$E(Y) = \int_0^1 y f_Y(y)\,dy = \int_0^1 12y^2(1 - y)^2\,dy = \frac{2}{5}$$

Therefore,

$$\begin{aligned}\text{Cov}(X, Y) &= E(XY) - E(X)E(Y) \\ &= \frac{2}{15} - \left(\frac{2}{5}\right)\left(\frac{2}{5}\right) \\ &= -\frac{2}{75}\end{aligned}$$

If $C_{ij} = 0$, the random variables X_i and X_j are said to be *uncorrelated*. It is easy to show that if X_i and X_j are independent random variables, then $C_{ij} = 0$ (see Prob. 4.8). In general, though, the converse is not true (see Prob. 4.9). However, if X_i and X_j are jointly normally distributed random variables with $C_{ij} = 0$, then they are also independent (see Prob. 4.10).

We now give two definitions that will shed some light on the significance of the covariance C_{ij}. If $C_{ij} > 0$, then X_i and X_j are said to be *positively correlated*. In this case, $X_i > \mu_i$ and $X_j > \mu_j$ tend to occur together, and $X_i < \mu_i$ and $X_j < \mu_j$ also tend to occur together [see Eq. (4.1)]. Thus, for positively correlated random variables, if one is large, the other is likely to be large also. If $C_{ij} < 0$, then X_i and X_j are said to be *negatively correlated*. In this case, $X_i > \mu_i$ and $X_j < \mu_j$ tend to occur together, and $X_i < \mu_i$ and $X_j > \mu_j$ also tend to occur together. Thus, for negatively correlated random variables, if one is large, the other is likely to be small. We give examples of positively and negatively correlated random variables in the next section.

If $X_1, X_2, \ldots, X_n$ are simulation output data (for example, X_i might be the delay D_i for the queueing example of Sec. 1.4), we shall often need to know not only the mean μ_i and variance σ_i^2 for $i = 1, 2, \ldots, n$, but also a measure of the dependence between X_i and X_j for $i \neq j$. However, the difficulty with using C_{ij} as a measure of dependence between X_i and X_j is that it is not dimensionless, which makes it interpretation troublesome. (If X_i and X_j are in units of minutes, say, then C_{ij} is in units of minutes squared.) As a result, we use the *correlation* ρ_{ij}, defined by

$$\rho_{ij} = \frac{C_{ij}}{\sqrt{\sigma_i^2 \sigma_j^2}} \qquad \begin{matrix} i = 1, 2, \ldots, n \\ j = 1, 2, \ldots, n \end{matrix} \tag{4.2}$$

as our primary measure of the (linear) dependence (see Prob. 4.11) between X_i and X_j. [We shall also denote the correlation between X_i and X_j by $\text{Cor}(X_i, X_j)$.] Since the denominator in Eq. (4.2) is positive, it is clear that ρ_{ij} has the same sign as C_{ij}. Furthermore, it can be shown that $-1 \leq \rho_{ij} \leq 1$ for all i and j (see Prob. 4.12). If ρ_{ij} is close to $+1$, then X_i and X_j are highly positively correlated. On the other hand, if ρ_{ij} is close to -1, then X_i and X_j are highly negatively correlated.

EXAMPLE 4.18. For the random variables in Example 4.11, it can be shown that $\text{Var}(X) = \text{Var}(Y) = \frac{1}{25}$. Therefore,

$$\text{Cor}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\,\text{Var}(Y)}} = \frac{-\frac{2}{75}}{\frac{1}{25}} = -\frac{2}{3}$$

4.3 SIMULATION OUTPUT DATA AND STOCHASTIC PROCESSES

Since most simulation models use random variables as input, the simulation output data are themselves random, and care must be taken in drawing conclusions about the model's true characteristics, e.g., the (expected) average delay in the queueing example of Sec. 1.4. In this and the next three sections we lay the groundwork for a careful treatment of output data analysis in Chaps. 9 and 10.

A *stochastic process* is a collection of "similar" random variables ordered over time, which are all defined on a common sample space. The set of all possible values that these random variables can take on is called the *state space*. If the collection is $X_1, X_2, \ldots$, then we have a *discrete-time* stochastic process. If the collection is $\{X(t), t \geq 0\}$, then we have a *continuous-time* stochastic process.

EXAMPLE 4.19. Consider a single-server queueing system, e.g., the *M/M/*1 queue, with IID interarrival times $A_1, A_2, \ldots$, IID service times $S_1, S_2, \ldots$, and customers served in a FIFO manner. Relative to the experiment of generating the random variates $A_1, A_2, \ldots$ and $S_1, S_2, \ldots$, one can define the discrete-time stochastic process of delays in queue $D_1, D_2, \ldots$ as follows (see Prob. 4.14):

$$D_1 = 0$$

$$D_{i+1} = \max\{D_i + S_i - A_{i+1}, 0\} \qquad \text{for } i = 1, 2, \ldots$$

Thus, the simulation maps the input random variables (i.e., the A_i's and the S_i's) into the output stochastic process $D_1, D_2, \ldots$ of interest. Here, the state space is the set of nonnegative real numbers. Note that D_i and D_{i+1} are positively correlated. (Why?)

EXAMPLE 4.20. For the queueing system of Example 4.19, let $Q(t)$ be the number of customers in the queue at time t. Then $\{Q(t), t \geq 0\}$ is a continuous-time stochastic process with state space $\{0, 1, 2, \ldots\}$.

EXAMPLE 4.21. For the inventory system of Sec. 1.5, let C_i be the total cost (i.e., the sum of the ordering, holding, and shortage costs) in month i. Then $C_1, C_2, \ldots$ is a discrete-time stochastic process with state space the nonnegative real numbers.

To draw inferences about an underlying stochastic process from a set of simulation output data, one must sometimes make assumptions about the stochastic process that may not be strictly true in practice. (Without such assumptions, however, statistical analysis of the output data may not be possible.) An example of this is to assume that a stochastic process is covariance-stationary, a property that we now define. A discrete-time stochastic process $X_1, X_2, \ldots$ is said to be *covariance-stationary* if

$$\mu_i = \mu \qquad \text{for } i = 1, 2, \ldots \text{ and } -\infty < \mu < \infty$$

$$\sigma_i^2 = \sigma^2 \qquad \text{for } i = 1, 2, \ldots \text{ and } \sigma^2 < \infty$$

and $C_{i,i+j} = \text{Cov}(X_i, X_{i+j})$ is independent of i for $j = 1, 2, \ldots$.

Thus, for a covariance-stationary process, the mean and variance are stationary over time (the common mean and variance are denoted by μ and σ^2, respectively),

and the covariance between two observations X_i and X_{i+j} depends only on the separation j (sometimes called the *lag*) and not on the actual time values i and $i + j$. (It is also possible to define a covariance-stationary continuous-time stochastic process $\{X(t), t \geq 0\}$ in an analogous way.)

For a covariance-stationary process, we denote the covariance and correlation between X_i and X_{i+j} by C_j and ρ_j, respectively, where

$$\rho_j = \frac{C_{i,i+j}}{\sqrt{\sigma_i^2 \sigma_{i+j}^2}} = \frac{C_j}{\sigma^2} = \frac{C_j}{C_0} \qquad \text{for } j = 0, 1, 2, \ldots$$

EXAMPLE 4.22. Consider the output process $D_1, D_2, \ldots$ for a covariance-stationary (see App. 4A for a discussion of this technical detail) $M/M/1$ queue with $\rho = \lambda/\omega < 1$ (recall that λ is the arrival rate and ω is the service rate). From results in Daley (1968), one can compute ρ_j, which we plot in Fig. 4.10 for $\rho = 0.5$ and 0.9. (Do not confuse ρ_j and ρ.) Note that the correlations ρ_j are positive and monotonically decrease to zero as j increases. In particular, $\rho_1 = 0.99$ for $\rho = 0.9$ and $\rho_1 = 0.78$ for $\rho = 0.5$. Furthermore, the convergence of ρ_j to zero is considerably slower for $\rho = 0.9$; in fact, ρ_{50} is (amazingly) 0.69. (In general, our experience indicates that output processes for queueing systems are positively correlated.)

EXAMPLE 4.23. Consider an (s, S) inventory system with zero delivery lag and backlogging. (For this inventory system, which is a simpler version of the one considered in Sec. 1.5, it is possible to compute the desired correlations analytically.) Let I_i, J_i, and Q_i denote, respectively, the amount of inventory on hand before ordering, the amount of inventory on hand after ordering, and the demand, each in month i. Assume that Q_i has a Poisson distribution (see Sec. 6.2.3 for further discussion) with a mean of 25; that is,

$$p(x) = P(Q_i = x) = \frac{e^{-25}(25)^x}{x!} \qquad \text{for } x = 0, 1, 2, \ldots$$

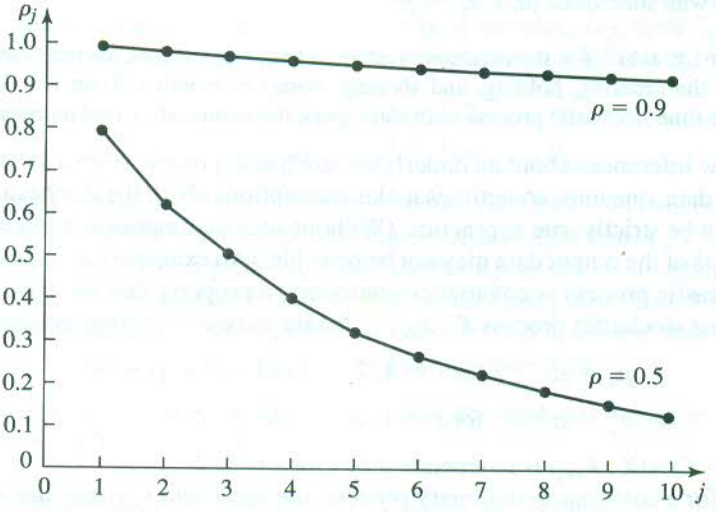

FIGURE 4.10
Correlation function ρ_j of the process $D_1, D_2, \ldots$ for the $M/M/1$ queue.

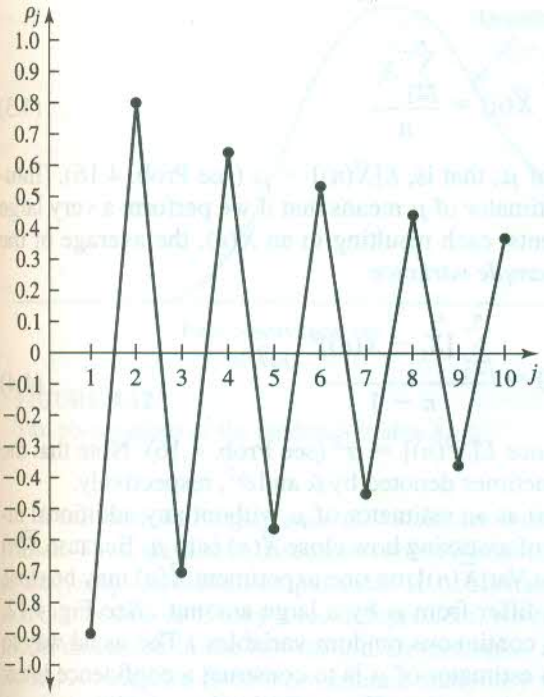

FIGURE 4.11
Correlation function ρ_j of the process $C_1, C_2, \ldots$ for an (s, S) inventory system.

If $I_i < s$, we order $S - I_i$ items ($J_i = S$) and incur a cost $K + i(S - I_i)$, where $K = 32$ and $i = 3$. If $I_i \geq s$, no order is placed ($J_i = I_i$) and no ordering cost is incurred. After J_i has been determined, the demand Q_i occurs. If $J_i - Q_i \geq 0$, a holding cost $h(J_i - Q_i)$ is incurred, where $h = 1$. If $J_i - Q_i < 0$, a shortage cost $\pi(Q_i - J_i)$ is incurred, where $\pi = 5$. In either case, $I_{i+1} = J_i - Q_i$. Let C_i be the total cost in month i, and assume that $s = 17$, $S = 57$, and $I_1 = S$. From results in Wagner (1969, p. A19), one can compute ρ_j for the output process $C_1, C_2, \ldots$, which we plot in Fig. 4.11. (See App. 4A for discussion of a technical detail.) Note that ρ_2 is positive, since for this particular system one tends to order every other month, incurring a large cost each time. On the other hand, ρ_1 is negative, because if one orders in a particular month (large cost), then it is likely that no order will be placed the next month (small cost).

If $X_1, X_2, \ldots$ is a stochastic process beginning at time 0 in a simulation, then it is quite likely *not* to be covariance-stationary (see App. 4A). However, for *some* simulations $X_{k+1}, X_{k+2}, \ldots$ will be approximately covariance-stationary if k is large enough, where k is the length of the *warmup period* (see Sec. 9.5.1).

4.4 ESTIMATION OF MEANS, VARIANCES, AND CORRELATIONS

Suppose that $X_1, X_2, \ldots, X_n$ are IID random variables (observations) with finite population mean μ and finite population variance σ^2 and that our primary objective is to estimate μ; the estimation of σ^2 is of secondary interest. Then the *sample*

mean

$$\overline{X}(n) = \frac{\sum_{i=1}^{n} X_i}{n} \tag{4.3}$$

is an unbiased (point) estimator of μ; that is, $E[\overline{X}(n)] = \mu$ (see Prob. 4.16). [Intuitively, $\overline{X}(n)$ being an unbiased estimator of μ means that if we perform a very large number of independent experiments, each resulting in an $\overline{X}(n)$, the average of the $\overline{X}(n)$'s will be μ.] Similarly, the *sample variance*

$$S^2(n) = \frac{\sum_{i=1}^{n} [X_i - \overline{X}(n)]^2}{n-1} \tag{4.4}$$

is an unbiased estimator of σ^2, since $E[S^2(n)] = \sigma^2$ (see Prob. 4.16). Note that the estimators $\overline{X}(n)$ and $S^2(n)$ are sometimes denoted by $\hat{\mu}$ and $\hat{\sigma}^2$, respectively.

The difficulty with using $\overline{X}(n)$ as an estimator of μ without any additional information is that we have no way of assessing how close $\overline{X}(n)$ is to μ. Because $\overline{X}(n)$ is a random variable with variance $\mathrm{Var}[\overline{X}(n)]$, on one experiment $\overline{X}(n)$ may be close to μ while on another $\overline{X}(n)$ may differ from μ by a large amount. (See Fig. 4.12, where the X_i's are assumed to be continuous random variables.) The usual way to assess the precision of $\overline{X}(n)$ as an estimator of μ is to construct a confidence interval for μ, which we discuss in the next section. However, the first step in constructing a confidence interval is to estimate $\mathrm{Var}[\overline{X}(n)]$. Since

$$\begin{aligned}
\mathrm{Var}[\overline{X}(n)] &= \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n^2}\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) \qquad \text{(because the } X_i\text{'s are independent)} \\
&= \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}
\end{aligned} \tag{4.5}$$

it is clear that, in general, the bigger the sample size n, the closer $\overline{X}(n)$ should be to μ (see Fig. 4.13). Furthermore, an unbiased estimator of $\mathrm{Var}[\overline{X}(n)]$ is obtained by replacing σ^2 in Eq. (4.5) by $S^2(n)$, resulting in

$$\widehat{\mathrm{Var}}[\overline{X}(n)] = \frac{S^2(n)}{n} = \frac{\sum_{i=1}^{n} [X_i - \overline{X}(n)]^2}{n(n-1)}$$

Observe that the expression for $\widehat{\mathrm{Var}}[\overline{X}(n)]$ has both an n and an $n-1$ in the denominator when it is rewritten in terms of the X_i's and $\overline{X}(n)$.

Finally, note that if the X_i's are independent, they are uncorrelated, and thus $\rho_j = 0$ for $j = 1, 2, \ldots, n-1$.

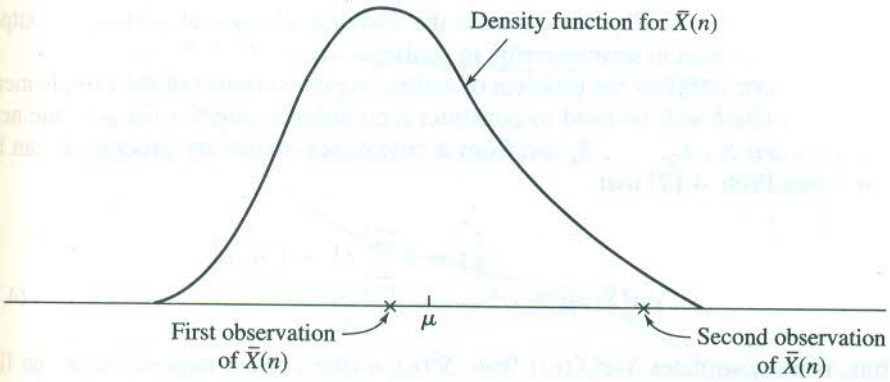

FIGURE 4.12
Two observations of the random variable $\bar{X}(n)$.

It has been our experience that simulation output data are almost always correlated. Thus, the above discussion about IID observations is not *directly* applicable to analyzing simulation output data. To understand the dangers of treating simulation output data as if they were independent, we shall use the covariance-stationary model discussed in the last section. In particular, assume that the random variables $X_1, X_2, \ldots, X_n$ are from a covariance-stationary stochastic process. Then it is still true that the sample mean $\bar{X}(n)$ is an unbiased estimator of μ; however, the sample variance $S^2(n)$ is no longer an unbiased estimator of σ^2. In fact, it can be shown [see Anderson (1994, p. 448)] that

$$E[S^2(n)] = \sigma^2\left[1 - 2\frac{\sum_{j=1}^{n-1}(1 - j/n)\rho_j}{n-1}\right] \tag{4.6}$$

Thus, if $\rho_j > 0$ (positive correlation), as is very often the case in practice, $S^2(n)$ will have a negative bias: $E[S^2(n)] < \sigma^2$. This is important because several simulation

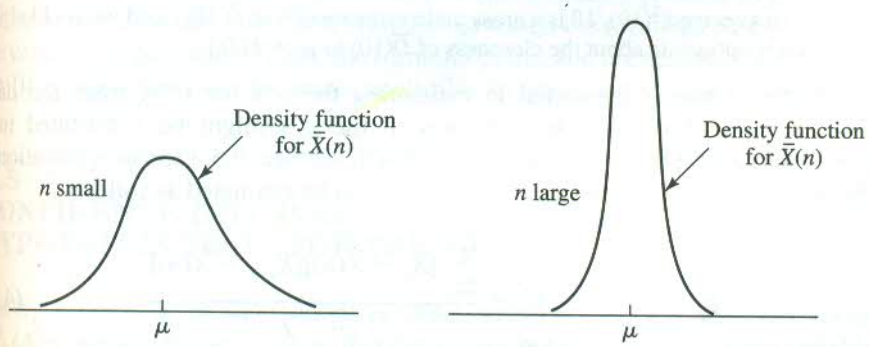

FIGURE 4.13
Distributions of $\bar{X}(n)$ for small and large n.

software products use $S^2(n)$ to estimate the variance of a set of simulation output data, which can lead to serious errors in analysis.

Let us now consider the problem of estimating the variance of the sample mean $\text{Var}[\overline{X}(n)]$ (which will be used to construct a confidence interval for μ in the next section) when $X_1, X_2, \ldots, X_n$ are from a covariance-stationary process. It can be shown (see Prob. 4.17) that

$$\text{Var}[\overline{X}(n)] = \sigma^2 \frac{\left[1 + 2\sum_{j=1}^{n-1} (1 - j/n)\rho_j\right]}{n} \tag{4.7}$$

Thus, if one estimates $\text{Var}[\overline{X}(n)]$ from $S^2(n)/n$ (the correct expression in the IID case), which has often been done historically, there are two sources of error: the bias in $S^2(n)$ as an estimator of σ^2 and the negligence of the correlation terms in Eq. (4.7). As a matter of fact, if we combine Eq. (4.6) and Eq. (4.7), we get

$$E\left[\frac{S^2(n)}{n}\right] = \frac{[n/a(n)] - 1}{n - 1} \text{Var}[\overline{X}(n)] \tag{4.8}$$

where $a(n)$ denotes the quantity in square brackets in Eq. (4.7). If $\rho_j > 0$, then $a(n) > 1$ and $E[S^2(n)/n] < \text{Var}[\overline{X}(n)]$.

EXAMPLE 4.24. Suppose that we have the data $D_1, D_2, \ldots, D_{10}$ from the process of delays $D_1, D_2, \ldots$ for a covariance-stationary $M/M/1$ queue with $\rho = 0.9$. Then, substituting the true correlations ρ_j (where $j = 1, 2, \ldots, 9$) into Eqs. (4.6) and (4.8), we get

$$E[S^2(10)] = 0.0328\sigma^2$$

and

$$E\left[\frac{S^2(10)}{10}\right] = 0.0034 \text{ Var}[\overline{D}(10)]$$

where

$$\sigma^2 = \text{Var}(D_i), \qquad \overline{D}(10) = \frac{\sum_{i=1}^{10} D_i}{10}, \qquad \text{and} \qquad S^2(10) = \frac{\sum_{i=1}^{10} [D_i - \overline{D}(10)]^2}{9}$$

Thus, on average $S^2(10)/10$ is a gross underestimate of $\text{Var}[\overline{D}(10)]$, and we are likely to be overly optimistic about the closeness of $\overline{D}(10)$ to $\mu = E(D_i)$.

Sometimes one is interested in estimating the ρ_j's (or C_j's) from the data $X_1, X_2, \ldots, X_n$. {For example, estimates of the ρ_j's might be substituted into Eq. (4.7) to obtain a better estimate of $\text{Var}[\overline{X}(n)]$; see Sec. 9.5.3 for an application.} If this is the case, ρ_j (for $j = 1, 2, \ldots, n - 1$) can be estimated as follows:

$$\hat{\rho}_j = \frac{\hat{C}_j}{S^2(n)}, \qquad \hat{C}_j = \frac{\sum_{i=1}^{n-j} [X_i - \overline{X}(n)][X_{i+j} - \overline{X}(n)]}{n - j} \tag{4.9}$$

[Other estimators of ρ_j are also used. For example, one could replace the $n - j$ in the denominator of $\hat{C}_j$ by n.] The difficulty with the estimator $\hat{\rho}_j$ (or any other

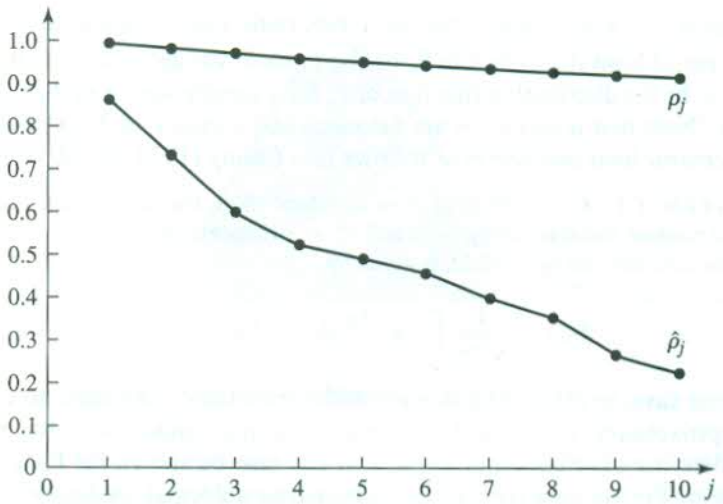

FIGURE 4.14
ρ_j and $\hat{\rho}_j$ of the process $D_1, D_2, \ldots$ for the $M/M/1$ queue with $\rho = 0.9$.

estimator of ρ_j) is that it is biased, it has a large variance unless n is very large, and it is correlated with other correlation estimators, that is, $\text{Cov}(\hat{\rho}_j, \hat{\rho}_k) \neq 0$. {In particular, $\hat{\rho}_{n-1}$ will be a poor estimator of ρ_{n-1} since it is based on the single product $[X_1 - \overline{X}(n)][X_n - \overline{X}(n)]$.} Thus, in general, "good" estimates of the ρ_j's will be difficult to obtain unless n is very large and j is small relative to n.

EXAMPLE 4.25. Suppose we have the data $D_1, D_2, \ldots, D_{100}$ from the process considered in Example 4.24. In Fig. 4.14 we plot $\hat{\rho}_j$ [as computed from Eq. (4.9)] and ρ_j for $j = 1, 2, \ldots, 10$. Note the poor quality of the correlation estimates.

Note that correlation estimates will not necessarily be zero when the X_i's are independent, since the estimator $\hat{\rho}_j$ is a random variable.

We have seen that simulation output data are correlated, and thus formulas from classical statistics based on IID observations cannot be used directly for estimating variances. However, we shall see in Chap. 9 that it is often possible to group simulation output data into new "observations" to which the formulas based on IID observations *can* be applied. Thus, the formulas in this and the next two sections based on IID observations are *indirectly* applicable to analyzing simulation output data.

4.5 CONFIDENCE INTERVALS AND HYPOTHESIS TESTS FOR THE MEAN

Let $X_1, X_2, \ldots, X_n$ be IID random variables with finite mean μ and finite variance σ^2. (Also assume that $\sigma^2 > 0$, so that the X_i's are not degenerate random variables.) In this section we discuss how to construct a confidence interval for μ and also the complementary problem of testing the hypothesis that $\mu = \mu_0$.

We begin with a statement of the most important result in probability theory, the classical central limit theorem. Let Z_n be the random variable $[\bar{X}(n) - \mu]/\sqrt{\sigma^2/n}$, and let $F_n(z)$ be the distribution function of Z_n for a sample size of n; that is , $F_n(z) = P(Z_n \leq z)$. [Note that μ and σ^2/n are the mean and variance of $\bar{X}(n)$, respectively.] Then the *central limit theorem* is as follows [see Chung (1974, p. 169) for a proof].

THEOREM 4.1. $F_n(z) \to \Phi(z)$ as $n \to \infty$, where $\Phi(z)$, the distribution function of a normal random variable with $\mu = 0$ and $\sigma^2 = 1$ (henceforth called a *standard normal random variable*; see Sec. 6.2.2), is given by

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-y^2/2}\, dy \qquad \text{for } -\infty < z < \infty$$

The theorem says, in effect, that if n is "sufficiently large," the random variable Z_n will be approximately distributed as a standard normal random variable, regardless of the underlying distribution of the X_i's. It can also be shown for large n that the sample mean $\bar{X}(n)$ is approximately distributed as a normal random variable with mean μ and variance σ^2/n.

The difficulty with using the above results in practice is that the variance σ^2 is generally unknown. However, since the sample variance $S^2(n)$ converges to σ^2 as n gets large, it can be shown that Theorem 4.1 remains true if we replace σ^2 by $S^2(n)$ in the expression for Z_n. With this change the theorem says that if n is sufficiently large, the random variable $t_n = [\bar{X}(n) - \mu]/\sqrt{S^2(n)/n}$ is approximately distributed as a standard normal random variable. It follows for large n that

$$P\left(-z_{1-\alpha/2} \leq \frac{\bar{X}(n) - \mu}{\sqrt{S^2(n)/n}} \leq z_{1-\alpha/2}\right)$$

$$= P\left[\bar{X}(n) - z_{1-\alpha/2}\sqrt{\frac{S^2(n)}{n}} \leq \mu \leq \bar{X}(n) + z_{1-\alpha/2}\sqrt{\frac{S^2(n)}{n}}\right]$$

$$\approx 1 - \alpha \tag{4.10}$$

where the symbol $\approx$ means "approximately equal" and $z_{1-\alpha/2}$ (for $0 < \alpha < 1$) is the upper $1 - \alpha/2$ critical point for a standard normal random variable (see Fig. 4.15 and the last line of Table T.1 of the Appendix at the back of the book). Therefore, if n is sufficiently large, an approximate $100(1 - \alpha)$ percent confidence interval for μ is given by

$$\bar{X}(n) \pm z_{1-\alpha/2}\sqrt{\frac{S^2(n)}{n}} \tag{4.11}$$

For a given set of data $X_1, X_2, \ldots, X_n$, the lower confidence-interval endpoint $l(n, \alpha) = \bar{X}(n) - z_{1-\alpha/2}\sqrt{S^2(n)/n}$ and the upper confidence-interval endpoint $u(n, \alpha) = \bar{X}(n) + z_{1-\alpha/2}\sqrt{S^2(n)/n}$ are just numbers (actually, specific realizations of random variables) and the confidence interval $[l(n, \alpha), u(n, \alpha)]$ either contains μ or does not contain μ. Thus, there is nothing probabilistic about the single confidence interval $[l(n, \alpha), u(n, \alpha)]$ *after* the data have been obtained and the interval's

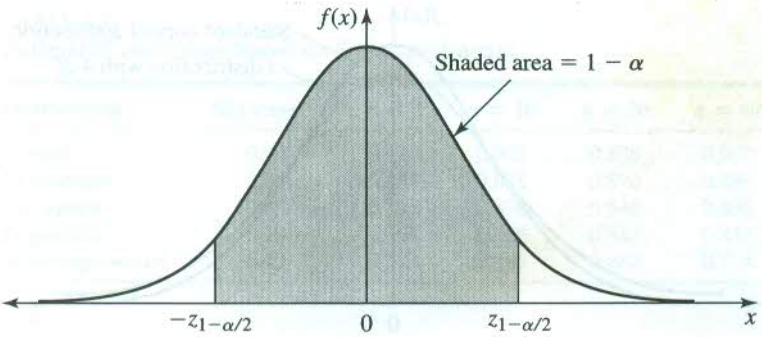

FIGURE 4.15
Density function for the standard normal distribution.

endpoints have been given numerical values. The correct interpretation to give to the confidence interval (4.11) is as follows [see (4.10)]: If one constructs a very large number of independent $100(1-\alpha)$ percent confidence intervals, each based on n observations, where n is sufficiently large, the proportion of these confidence intervals that contain (cover) μ should be $1-\alpha$. We call this proportion the *coverage* for the confidence interval.

The difficulty in using (4.11) to construct a confidence interval for μ is in knowing what "n sufficiently large" means. It turns out that the more skewed (i.e., nonsymmetric) the underlying distribution of the X_i's, the larger the value of n needed for the distribution of t_n to be closely approximated by $\Phi(z)$. (See the discussion later in this section.) If n is chosen too small, the actual coverage of a desired $100(1-\alpha)$ percent confidence interval will generally be less than $1-\alpha$. This is why the confidence interval given by (4.11) is stated to be only approximate.

In light of the above discussion, we now develop an alternative confidence-interval expression. If the X_i's are *normal* random variables, the random variable $t_n = [\overline{X}(n) - \mu]/\sqrt{S^2(n)/n}$ has a t distribution with $n-1$ degrees of freedom (df) [see, for example, Hogg and Craig (1995, pp. 181–182)], and an *exact* (for any $n \geq 2$) $100(1-\alpha)$ percent confidence interval for μ is given by

$$\overline{X}(n) \pm t_{n-1,1-\alpha/2}\sqrt{\frac{S^2(n)}{n}} \tag{4.12}$$

where $t_{n-1,1-\alpha/2}$ is the upper $1-\alpha/2$ critical point for the t distribution with $n-1$ df. These critical points are given in Table T.1 of the Appendix at the back of the book. Plots of the density functions for the t distribution with 4 df and for the standard normal distribution are given in Fig. 4.16. Note that the t distribution is less peaked and has longer tails than the normal distribution, so, for any finite n, $t_{n-1,1-\alpha/2} > z_{1-\alpha/2}$. We call (4.12) the *t confidence interval.*

The quantity that we add to and subtract from $\overline{X}(n)$ in (4.12) to construct the confidence interval is called the *half-length* of the confidence interval. It is a measure of how precisely we know μ. It can be shown that if we increase the sample size from n to $4n$ in (4.12), then the half-length is decreased by a factor of approximately 2 (see Prob. 4.20).

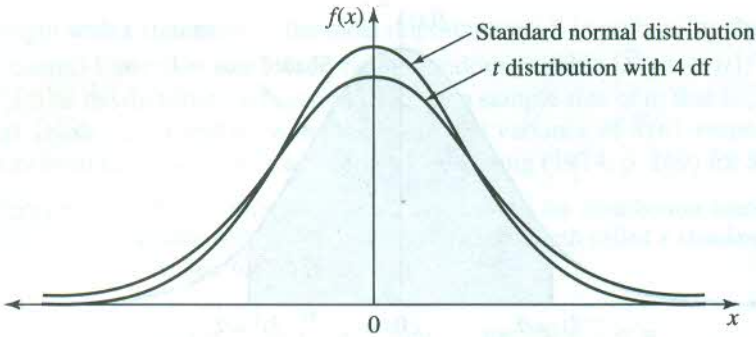

FIGURE 4.16
Density functions for the t distribution with 4 df and for the standard normal distribution.

In practice, the distribution of the X_i's will rarely be normal, and the confidence interval given by (4.12) will also be approximate in terms of coverage. Since $t_{n-1,1-\alpha/2} > z_{1-\alpha/2}$, the confidence interval given by (4.12) will be larger than the one given by (4.11) and will generally have coverage closer to the desired level $1 - \alpha$. For this reason, we recommend using (4.12) to construct a confidence interval for μ. Note that $t_{n-1,1-\alpha/2} \to z_{1-\alpha/2}$ as $n \to \infty$; in particular, $t_{40,0.95}$ differs from $z_{0.95}$ by less than 3 percent. However, in most of our applications of (4.12) in Chaps. 9, 10, and 12, n will be small enough for the difference between (4.11) and (4.12) to be appreciable.

EXAMPLE 4.26. Suppose that the 10 observations 1.20, 1.50, 1.68, 1.89, 0.95, 1.49, 1.58, 1.55, 0.50, and 1.09 are from a normal distribution with unknown mean μ and that our objective is to construct a 90 percent confidence interval for μ. From these data we get

$$\overline{X}(10) = 1.34 \quad \text{and} \quad S^2(10) = 0.17$$

which results in the following confidence interval for μ:

$$\overline{X}(10) \pm t_{9,0.95}\sqrt{\frac{S^2(10)}{10}} = 1.34 \pm 1.83\sqrt{\frac{0.17}{10}} = 1.34 \pm 0.24$$

Note that (4.12) was used to construct the confidence interval and that $t_{9,0.95}$ was taken from Table T.1. Therefore, subject to the interpretation stated above, we claim with 90 percent confidence that μ is in the interval [1.10, 1.58].

We now discuss how the coverage of the confidence interval given by (4.12) is affected by the distribution of the X_i's. In Table 4.1 we give estimated coverages for 90 percent confidence intervals based on 500 independent experiments for each of the sample sizes $n = 5$, 10, 20, and 40 and each of the distributions normal, exponential, chi square with 1 df (a standard normal random variable squared; see the discussion of the gamma distribution in Sec. 6.2.2), lognormal (e^Y, where Y is a standard normal random variable; see Sec. 6.2.2), and hyperexponential whose distribution function is given by

$$F(x) = 0.9F_1(x) + 0.1\,F_2(x)$$

TABLE 4.1
Estimated coverages based on 500 experiments

Distribution	Skewness ν	$n = 5$	$n = 10$	$n = 20$	$n = 40$
Normal	0.00	0.910	0.902	0.898	0.900
Exponential	2.00	0.854	0.878	0.870	0.890
Chi square	2.83	0.810	0.830	0.848	0.890
Lognormal	6.18	0.758	0.768	0.842	0.852
Hyperexponential	6.43	0.584	0.586	0.682	0.774

where $F_1(x)$ and $F_2(x)$ are the distribution functions of exponential random variables with means 0.5 and 5.5, respectively. For example, the table entry for the exponential distribution and $n = 10$ was obtained as follows. Ten observations were generated from an exponential distribution with a *known* mean μ, a 90 percent confidence interval was constructed using (4.12), and it was determined whether the interval contained μ. (This constituted one experiment.) Then the whole procedure was repeated 500 times, and 0.878 is the proportion of the 500 confidence intervals that contained μ. Note that the coverage for the normal distribution and $n = 10$ is 0.902 rather than the expected 0.900, since the table is based on 500 rather than an infinite number of experiments.

Observe from the table that for a particular distribution, coverage generally gets closer to 0.90 as n gets larger, which follows from the central limit theorem (see Prob. 4.22). (The results for the exponential distribution would also probably follow this behavior if the number of experiments were larger.) Notice also that for a particular n, coverage decreases as the skewness of the distribution gets larger, where skewness is defined by

$$\nu = \frac{E[(X - \mu)^3]}{(\sigma^2)^{3/2}} \qquad -\infty < \nu < \infty$$

The skewness, which is a measure of symmetry, is equal to 0 for a symmetric distribution such as the normal. We conclude from the table that the larger the skewness of the distribution in question, the larger the sample size needed to obtain satisfactory (close to 0.90) coverage.

Assume that $X_1, X_2, \ldots, X_n$ are normally distributed (or are approximately so) and that we would like to test the null hypothesis H_0 that $\mu = \mu_0$, where μ_0 is a fixed, hypothesized value for μ. Intuitively, we would expect that if $|\overline{X}(n) - \mu_0|$ is large [recall that $\overline{X}(n)$ is the point estimator for μ], H_0 is not likely to be true. However, to develop a test with known statistical properties, we need a statistic (a function of the X_i's) whose distribution is known when H_0 is true. It follows from the above discussion that if H_0 is true, the statistic $t_n = [\overline{X}(n) - \mu_0]/\sqrt{S^2(n)/n}$ will have a t distribution with $n - 1$ df. Therefore, consistent with our intuitive discussion above, the form of our (two-tailed) hypothesis test for $\mu = \mu_0$ is

$$\text{If} \quad |t_n| \begin{cases} > t_{n-1,1-\alpha/2} & \text{reject } H_0 \\ \leq t_{n-1,1-\alpha/2} & \text{"accept" } H_0 \end{cases} \tag{4.13}$$

The portion of the real line that corresponds to rejection of H_0, namely, the set of all x such that $|x| > t_{n-1,1-\alpha/2}$, is called the *critical region* for the test, and the probability that the statistic t_n falls in the critical region given that H_0 is true, which is clearly equal to α, is called the *level* (or size) of the test. Typically, an experimenter will choose the level equal to 0.05 or 0.10. We call the hypothesis test given by (4.13) the *t test.*

When one performs a hypothesis test, two types of errors can be made. If one rejects H_0 when in fact it is true, this is called a *Type I error.* The probability of a Type I error is equal to the level α and is thus under the experimenter's control. If one accepts H_0 when it is false, this is called a *Type II error.* For a fixed level α and sample size n, the probability of a *Type II error*, which we denote by β, depends on what is actually true (as compared to H_0) and may be unknown. We call $\delta = 1 - \beta$ the *power* of the test, and it is equal to the probability of rejecting H_0 when it is false. (Clearly, a test with high power is desirable.) If α is fixed, the power of a test can be increased only by increasing n. Since the power of a test may be low and unknown to us, we shall henceforth say that we "fail to reject H_0" (instead of "accept H_0") when the statistic t_n does not lie in the critical region. (When H_0 is not rejected, we generally do not know with any certainty whether H_0 is true or whether H_0 is false, since our test might not be powerful enough to detect any difference between H_0 and what is actually true.)

EXAMPLE 4.27. For the data of Example 4.26, suppose that we would like to test the null hypothesis H_0 that $\mu = 1$ at level $\alpha = 0.10$. Since

$$t_{10} = \frac{\bar{X}(10) - 1}{\sqrt{S^2(10)/10}} = \frac{0.34}{\sqrt{0.17/10}} = 2.65 > 1.83 = t_{9,0.95}$$

we reject H_0.

EXAMPLE 4.28. For the null hypothesis H_0 that $\mu = 1$ in Example 4.27, we can estimate the power of the test when, in fact, the X_i's have a normal distribution with mean $\mu = 1.5$ and standard deviation $\sigma = 1$. We randomly generated 1000 independent observations of the statistic $t_{10} = [\bar{X}(10) - 1]/\sqrt{S^2(10)/10}$ under the assumption that $\mu = 1.5$ and $\sigma = 1$ (the X_i's were, of course, normal). For 447 out of the 1000 observations, $|t_{10}| > 1.83$ and, therefore, the estimated power is $\hat{\delta} = 0.447$. Thus, if $\mu = 1.5$ and $\sigma = 1$, we will only reject the null hypothesis $\mu = 1$ approximately 45 percent of the time for a test at level $\alpha = 0.10$. To see what effect the standard deviation σ has on the power of the test, we generated 1000 observations of t_{10} when $\mu = 1.5$ and $\sigma = 0.75$ and also 1000 observations of t_{10} when $\mu = 1.5$ and $\sigma = 0.5$ (all X_i's were normal). The estimated powers were $\hat{\delta} = 0.619$ and $\hat{\delta} = 0.900$, respectively. It is not surprising that the power is apparently a decreasing function of σ, since we would expect to distinguish better between the true mean 1.5 and the hypothesized mean 1 when σ is small. [Note that in the case of normal sampling, as in this example, the power of the test can actually be computed exactly, obviating the need for simulation as done here; see advanced texts on statistics such as Bickel and Doksum (1977) with reference to the *noncentral t* distribution.]

It should be mentioned that there is an intimate relationship between the confidence interval given by (4.12) and the hypothesis test given by (4.13). In particular,

rejection of the null hypothesis H_0 that $\mu = \mu_0$ is equivalent to μ_0 not being contained in the confidence interval for μ, assuming the same value of α for both the hypothesis test and the confidence interval.

4.6 THE STRONG LAW OF LARGE NUMBERS

The second most important result in probability theory (after the central limit theorem) is arguably the strong law of large numbers. Let $X_1, X_2, \ldots, X_n$ be IID random variables with finite mean μ. Then the *strong law of large numbers* is as follows [see Chung (1974, p. 126) for a proof].

THEOREM 4.2. $\overline{X}(n) \to \mu$ w.p. 1 as $n \to \infty$.

The theorem says, in effect, that if one performs an infinite number of experiments, each resulting in an $\overline{X}(n)$, and n is sufficiently large, then $\overline{X}(n)$ will be arbitrarily close to μ for almost all the experiments.

EXAMPLE 4.29. Suppose that $X_1, X_2, \ldots$ are IID normal random variables with $\mu = 1$ and $\sigma^2 = 0.01$. Figure 4.17 plots the values of $\overline{X}(n)$ for various n that resulted from sampling from this distribution. Note that $\overline{X}(n)$ differed from μ by less than 1 percent for $n \geq 28$.

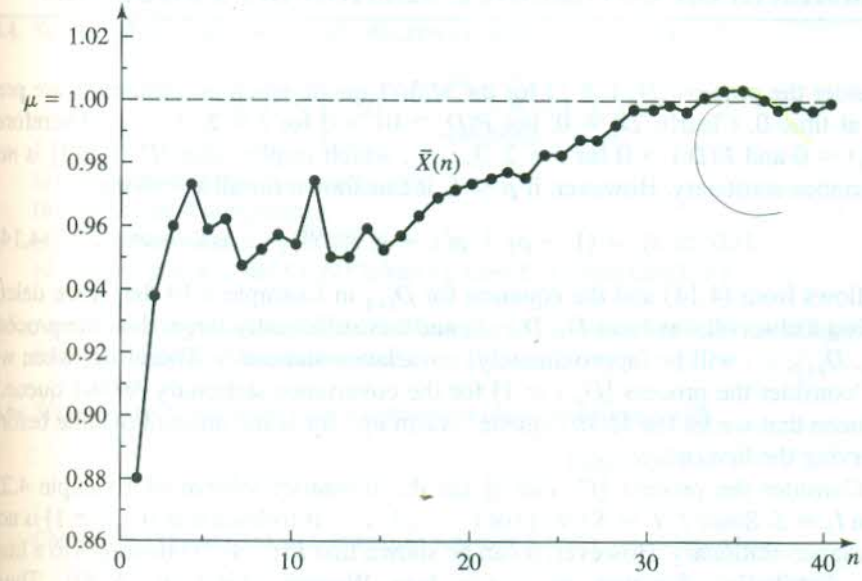

FIGURE 4.17
$\overline{X}(n)$ for various values of n when the X_i's are normal random variables with $\mu = 1$ and $\sigma^2 = 0.01$.

4.7 THE DANGER OF REPLACING A PROBABILITY DISTRIBUTION BY ITS MEAN

Simulation analysts have sometimes replaced an input probability distribution by its mean in their simulation models. This practice may be caused by a lack of understanding on the part of the analyst or by lack of information on the actual form of the distribution (e.g., only an estimate of the mean of the distribution is available). The following example illustrates the danger of this practice.

EXAMPLE 4.30. Consider a manufacturing system consisting of a single machine tool. Suppose that "raw" parts arrive to the machine with exponential interarrival times having a mean of 1 minute and that processing times at the machine are exponentially distributed with a mean of 0.99 minute. Thus, this system is an *M/M/*1 queue with utilization factor $\rho = 0.99$. Furthermore, it can be shown that the average delay in queue of a part in the long run is 98.01 minutes [see App. 1B or Gross and Harris (1998, p. 67)]. On the other hand, if we replace each distribution by its corresponding mean (i.e., if customers arrive at times 1 minute, 2 minutes, . . . and if each part has a processing time of exactly 0.99 minute), then no part is ever delayed in the queue. In general, the variances as well as the means of the input distributions affect the output measures for queueing-type systems, as noted at the end of App. 1B.

APPENDIX 4A COMMENTS ON COVARIANCE-STATIONARY PROCESSES

Consider the process $\{D_i, i \geq 1\}$ for the *M/M/*1 queue when no customers are present at time 0. Clearly, $D_1 = 0$, but $P(D_i > 0) > 0$ for $i = 2, 3, \ldots$. Therefore, $E(D_1) = 0$ and $E(D_i) > 0$ for $i = 2, 3, \ldots$, which implies that $\{D_i, i \geq 1\}$ is *not* covariance-stationary. However, if $\rho < 1$, it can shown for all $x \geq 0$ that

$$P(D_i \leq x) \rightarrow (1 - \rho) + \rho(1 - e^{-(\omega - \lambda)x}) \qquad \text{as } i \rightarrow \infty \tag{4.14}$$

It follows from (4.14) and the equation for D_{i+1} in Example 4.19 that if we delete the first k observations from $D_1, D_2, \ldots$ and k is sufficiently large, then the process $D_{k+1}, D_{k+2}, \ldots$ will be (approximately) covariance-stationary. Therefore, when we say "consider the process $\{D_i, i \geq 1\}$ for the covariance-stationary *M/M/*1 queue," we mean that we let the *M/M/*1 queue "warm up" for some amount of time before observing the first delay.

Consider the process $\{C_i, i \geq 1\}$ for the inventory system of Example 4.23 when $I_1 = S$. Since $P(I_i = S) \neq 1$ for $i = 2, 3, \ldots$, it follows that $\{C_i, i \geq 1\}$ is not covariance-stationary. However, it can be shown that $P(C_i \leq x)$ coverages to a limiting distribution function as $i \rightarrow \infty$ [see Wagner (1969, p. A48)]. Thus, $C_{k+1}, C_{k+2}, \ldots$ will be (approximately) covariance-stationary for k large. Furthermore, the correlations plotted in Fig. 4.11 are for an inventory system warmed up for some amount of time before the first cost is observed.