5

Statistical Models in Simulation

In modeling real-world phenomena, there are few situations where the actions of the entities within the system under study can be predicted completely. The world the model-builder sees is probabilistic rather than deterministic. There are many causes of variation. The time it takes a repairperson to fix a broken machine is a function of the complexity of the breakdown, whether the repairperson brought the proper replacement parts and tools to the site, whether another repairperson asks for assistance during the course of the repair, whether the machine operator receives a lesson in preventive maintenance, and so on. To the model-builder, these variations appear to occur by chance and cannot be predicted. However, some statistical model might well describe the time to make a repair.

An appropriate model can be developed by sampling the phenomenon of interest. Then, through educated guesses (or using software for the purpose), the model-builder would select a known distribution form, make an estimate of the parameter(s) of this distribution, and then test to see how good a fit has been obtained. Through continued efforts in the selection of an appropriate distribution form, a postulated model could be accepted. This multistep process is described in Chapter 9.

Section 5.1 contains a review of probability terminology and concepts. Some typical applications of statistical models, or distribution forms, are given in Section 5.2. Then, a number of selected discrete and continuous distributions are discussed in Sections 5.3 and 5.4. The selected distributions are those that describe a wide variety of probabilistic events and, further, appear in different contexts in other chapters of this text. Additional discussion about the distribution forms appearing in this chapter, and about distribution forms mentioned but not described, is available from a number of sources [Hines and Montgomery, 1990; Ross, 2002; Papoulis, 1990; Devore, 1999; Walpole and Myers, 2002; Law and Kelton, 2000]. Section 5.5 describes the Poisson process and its relationship to the exponential distribution. Section 5.6 discusses empirical distributions.

5.1 REVIEW OF TERMINOLOGY AND CONCEPTS

1. *Discrete random variables.* Let X be a random variable. If the number of possible values of X is finite, or countably infinite, X is called a discrete random variable. The possible values of X may be listed as $x_1, x_2, \ldots$ In the finite case, the list terminates; in the countably infinite case, the list continues indefinitely.

Example 5.1

The number of jobs arriving each week at a job shop is observed. The random variable of interest is X, where

$$X = \text{number of jobs arriving each week}$$

The possible values of X are given by the range space of X, which is denoted by R_X. Here $R_X = \{0, 1, 2, \ldots\}$.

Let X be a discrete random variable. With each possible outcome x_i in R_X, a number $p(x_i) = P(X = x_i)$ gives the probability that the random variable equals the value of x_i. The numbers $p(x_i)$, $i = 1, 2, \ldots$, must satisfy the following two conditions:

1. $p(x_i) \geq 0$, for all i
2. $\sum_{i=1}^{\infty} p(x_i) = 1$

The collection of pairs $(x_i, p(x_i))$, $i = 1, 2, \ldots$ is called the probability distribution of X, and $p(x_i)$ is called the probability mass function (pmf) of X.

Example 5.2

Consider the experiment of tossing a single die. Define X as the number of spots on the up face of the die after a toss. Then $R_X = \{1, 2, 3, 4, 5, 6\}$. Assume the die is loaded so that the probability that a given face lands up is proportional to the number of spots showing. The discrete probability distribution for this random experiment is given by

x_i	1	2	3	4	5	6
$p(x_i)$	1/21	2/21	3/21	4/21	5/21	6/21

The conditions stated earlier are satisfied—that is, $p(x_i) \geq 0$ for $i = 1, 2, \ldots, 6$ and $\sum_{i=1}^{\infty} p(x_i) = 1/21 + \cdots + 6/21 = 1$. The distribution is shown graphically in Figure 5.1.

2. *Continuous random variables.* If the range space R_X of the random variable X is an interval or a collection of intervals, X is called a continuous random variable. For a continuous random variable X, the probability that X lies in the interval $[a, b]$ is given by

$$P(a \leq X \leq b) = \int_a^b f(x)\,dx \tag{5.1}$$

The function $f(x)$ is called the probability density function (pdf) of the random variable X. The pdf satisfies the following conditions:

a. $f(x) \geq 0$ for all x in R_X

b. $\int_{R_X} f(x)dx = 1$

c. $f(x) = 0$ if x is not in R_X

As a result of Equation (5.1), for any specified value x_0, $P(X = x_0) = 0$, because

$$\int_{x_0}^{x_0} f(x)\,dx = 0$$

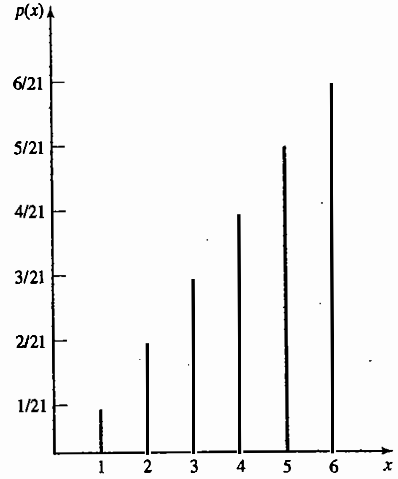

Figure 5.1 Probability mass function for loaded-die example.

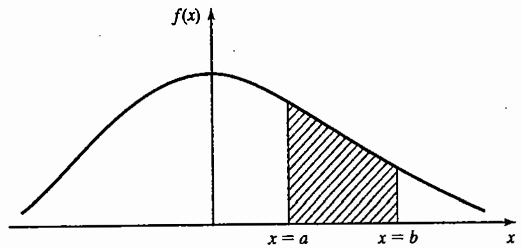

Figure 5.2 Graphical interpretation of $P(a < X < b)$.

$P(X = x_0) = 0$ also means that the following equations hold:

$$P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b) \tag{5.2}$$

The graphical interpretation of Equation (5.1) is shown in Figure 5.2. The shaded area represents the probability that X lies in the interval $[a, b]$.

Example 5.3

The life of a device used to inspect cracks in aircraft wings is given by X, a continuous random variable assuming all values in the range $x \geq 0$. The pdf of the lifetime, in years, is as follows:

$$f(x) = \begin{cases} \frac{1}{2} e^{-x/2}, & x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

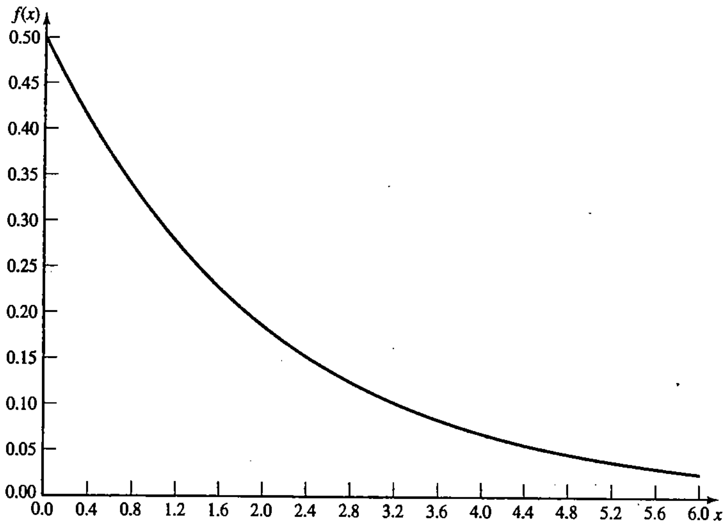

Figure 5.3 pdf for inspection-device life.

This pdf is shown graphically in Figure 5.3. The random variable X is said to have an exponential distribution with mean 2 years.

The probability that the life of the device is between 2 and 3 years is calculated as

$$P(2 \le X \le 3) = \frac{1}{2}\int_2^3 e^{-x/2}dx$$
$$= -e^{-3/2} + e^{-1} = -0.223 + 0.368 = 0.145$$

3. *Cumulative distribution function.* The cumulative distribution function (cdf), denoted by $F(x)$, measures the probability that the random variable X assumes a value less than or equal to x, that is, $F(x) = P(X \le x)$.

If X is discrete, then

$$F(x) = \sum_{\substack{\text{all} \\ x_i \le x}} p(x_i) \tag{5.3}$$

If X is continuous, then

$$F(x) = \int_{-\infty}^{x} f(t)dt \tag{5.4}$$

Some properties of the cdf are listed here:

a. F is a nondecreasing function. If $a < b$, then $F(a) \le F(b)$.
b. $\lim_{x \to \infty} F(x) = 1$
c. $\lim_{x \to -\infty} F(x) = 0$

All probability questions about X can be answered in terms of the cdf. For example,

$$P(a < X \le b) = F(b) - F(a) \quad \text{for all } a < b \tag{5.5}$$

For continuous distributions, not only does Equation (5.5) hold, but also the probabilities in Equation (5.2) are equal to $F(b) - F(a)$.

Example 5.4

The die-tossing experiment described in Example 5.2 has a cdf given as follows:

x	$(-\infty, 1)$	$[1, 2)$	$[2, 3)$	$[3, 4)$	$[4, 5)$	$[5, 6)$	$[6, \infty)$
$F(x)$	0	1/21	3/21	6/21	10/21	15/21	21/21

where $[a, b) = \{a \le x < b\}$. The cdf for this example is shown graphically in Figure 5.4.

If X is a discrete random variable with possible values $x_1, x_2, \ldots$, where $x_1 < x_2 < \ldots$, the cdf is a step function. The value of the cdf is constant in the interval $[x_{i-1}, x_i)$ and then takes a step, or jump, of size $p(x_i)$ at x_i. Thus, in Example 5.4, $p(3) = 3/21$, which is the size of the step when $x = 3$.

Example 5.5

The cdf for the device described in Example 5.3 is given by

$$F(x) = \frac{1}{2}\int_0^x e^{-t/2}dt = 1 - e^{-x/2}$$

The probability that the device will last for less than 2 years is given by

$$P(0 \le X \le 2) = F(2) - F(0) = F(2) = 1 - e^{-1} = 0.632$$

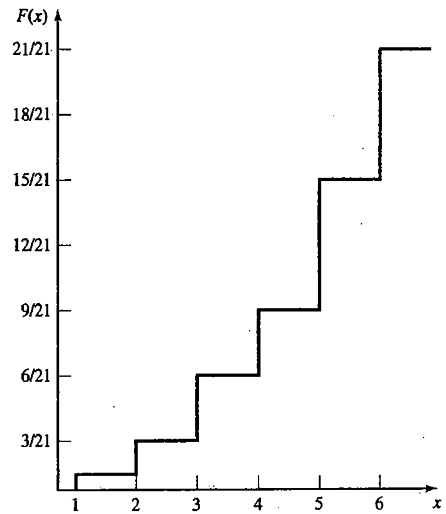

Figure 5.4 cdf for loaded-die example.

The probability that the life of the device is between 2 and 3 years is calculated as

$$P(2 \le X \le 3) = F(3) - F(2) = (1-e^{-3/2}) - (1-e^{-1})$$
$$= -e^{-3/2} + e^{-1} = -0.223 + 0.368 = 0.145$$

as found in Example 5.3.

4. *Expectation.* An important concept in probability theory is that of the expectation of a random variable. If X is a random variable, the expected value of X, denoted by $E(X)$, for discrete and continuous variables is defined as follows:

$$E(X) = \sum_{\text{all } i} x_i p(x_i) \qquad \text{if } X \text{ is discrete} \tag{5.6}$$

and

$$E(X) = \int_{-\infty}^{\infty} x f(x)dx \qquad \text{if } X \text{ is continuous} \tag{5.7}$$

The expected value $E(X)$ of a random variable X is also referred to as the mean, μ, or the first moment of X. The quantity $E(X^n)$, $n \ge 1$, is called the nth moment of X, and is computed as follows:

$$E(X^n) = \sum_{\text{all } i} x_i^n p(x_i) \qquad \text{if } X \text{ is discrete} \tag{5.8}$$

and

$$E(X^n) = \int_{-\infty}^{\infty} x^n f(x)dx \qquad \text{if } X \text{ is continuous} \tag{5.9}$$

The variance of a random variable, X, denoted by $V(X)$ or $\text{var}(X)$ or σ^2, is defined by

$$V(X) = E[(X - E[X])^2]$$

A useful identity in computing $V(X)$ is given by

$$V(X) = E(X^2) - [E(X)]^2 \tag{5.10}$$

The mean $E(X)$ is a measure of the central tendency of a random variable. The variance of X measures the expected value of the squared difference between the random variable and its mean. Thus, the variance, $V(X)$, is a measure of the spread or variation of the possible values of X around the mean $E(X)$. The standard deviation, σ, is defined to be the square root of the variance, σ^2. The mean, $E(X)$, and the standard deviation, $\sigma = \sqrt{V(X)}$, are expressed in the same units.

Example 5.6

The mean and variance of the die-tossing experiment described in Example 5.2 are computed as follows:

$$E(X) = 1\left(\frac{1}{21}\right) + 2\left(\frac{2}{21}\right) + \cdots + 6\left(\frac{6}{21}\right) = \frac{91}{21} = 4.33$$

To compute $V(X)$ from Equation (5.10), first compute $E(X^2)$ from Equation (5.8) as follows:

$$E(X^2) = 1^2\left(\frac{1}{21}\right) + 2^2\left(\frac{2}{21}\right) + \cdots + 6^2\left(\frac{6}{21}\right) = 21$$

Thus,

$$V(X) = 21 - \left(\frac{91}{21}\right)^2 = 21 - 18.78 = 2.22$$

and

$$\sigma = \sqrt{V(X)} = 1.49$$

Example 5.7

The mean and variance of the life of the device described in Example 5.3 are computed as follows:

$$E(X) = \frac{1}{2}\int_0^\infty xe^{-x/2}dx = -xe^{-x/2}\Big|_0^\infty + \int_0^\infty e^{-x/2}dx$$
$$= 0 + \frac{1}{1/2}e^{-x/2}\Big|_0^\infty = 2 \text{ years}$$

To compute $V(X)$ from Equation (5.10), first compute $E(X^2)$ from Equation (5.9) as follows:

$$E(X^2) = \frac{1}{2}\int_0^\infty x^2 e^{-x/2}dx$$

Thus,

$$E(X^2) = -x^2 e^{-x/2}\Big|_0^\infty + 2\int_0^\infty xe^{-x/2}dx = 8$$

giving

$$V(X) = 8 - 2^2 = 4 \text{ years}^2$$

and

$$\sigma = \sqrt{V(X)} = 2 \text{ years}$$

With a mean life of 2 years and a standard deviation of 2 years, most analysts would conclude that actual lifetimes, X, have a fairly large variability.

5. *The mode.* The mode is used in describing several statistical models that appear in this chapter. In the discrete case, the mode is the value of the random variable that occurs most frequently. In the continuous case, the mode is the value at which the pdf is maximized. The mode might not be unique; if the modal value occurs at two values of the random variable, the distribution is said to be bimodal.

5.2 USEFUL STATISTICAL MODELS

Numerous situations arise in the conduct of a simulation where an investigator may choose to introduce probabilistic events. In Chapter 2, queueing, inventory, and reliability examples were given. In a queueing system, interarrival and service times are often probabilistic. In an inventory model, the time between demands and the lead times (time between placing and receiving an order) can be probabilistic. In a reliability model, the time to failure could be probabilistic. In each of these instances, the simulation analyst desires to generate random events and to use a known statistical model if the underlying distribution can be found. In the following paragraphs, statistical models appropriate to these application areas will be discussed. Additionally, statistical models useful in the case of limited data are mentioned.

1. *Queueing systems.* In Chapter 2, examples of waiting-line problems were given. In Chapters 2, 3, and 4, these problems were solved via simulation. In the queueing examples, interarrival- and service-time patterns were given. In these examples, the times between arrivals and the service times were always probabilistic, as is usually the case. However, it is possible to have a constant interarrival time (as in the case of a line moving at a constant speed in the assembly of an automobile), or a constant service time (as in the case of robotized spot welding on the same assembly line). The following example illustrates how probabilistic interarrival times might occur.

Example 5.8

Mechanics arrive at a centralized tool crib as shown in Table 5.1. Attendants check in and check out the requested tools to the mechanics. The collection of data begins at 10:00 A.M. and continues until 20 different interarrival times are recorded. Rather than record the actual time of day, the absolute time from a given origin could have been computed. Thus, the first mechanic could have arrived at time zero, the second mechanic at time 7:13 (7 minutes, 13 seconds), and so on.

Example 5.9

Another way of presenting interarrival data is to find the number of arrivals per time period. Here, such arrivals occur over approximately 1 1/2 hours; it is convenient to look at 10-minute time intervals for the first 20 mechanics. That is, in the first 10-minute time period, one arrival occurred at 10:05::03. In the second time period, two mechanics arrived, and so on. The results are summarized in Table 5.2. This data could then be plotted in a histogram, as shown in Figure 5.5.

Table 5.1 Arrival Data

Arrival Number	*Arrival (Hour:Minutes::Seconds)*	*Interarrival Time (Minutes::Seconds)*
1	10:05::03	—
2	10:12::16	7::13
3	10:15::48	3::32
4	10:24::27	8::39
5	10:32::19	7::52
6	10:35::43	3::24
7	10:39::51	4::08
8	10:40::30	0::39
9	10:41::17	0::47
10	10:44::12	2::55
11	10:45::47	1::35
12	10:50::47	5::00
13	11:00::05	9::18
14	11:04::58	4::53
15	11:06::12	1::14
16	11:11::23	5::11
17	11:16::31	5::08
18	11:17::18	0::47
19	11:21::26	4::08
20	11:24::43	3::17
21	11:31::19	6::36

Table 5.2 Arrivals in Successive Time Periods

Time Period	*Number of Arrivals*	*Time Period*	*Number of Arrivals*
1	1	6	1
2	2	7	3
3	1	8	3
4	3	9	2
5	4	—	—

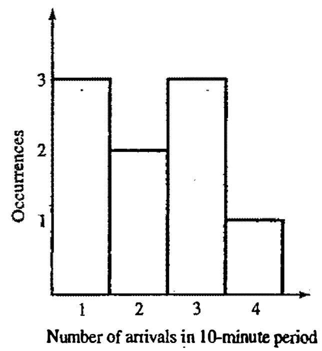

Figure 5.5 Histogram of arrivals per time period.

The distribution of time between arrivals and the distribution of the number of arrivals per time period are important in the simulation of waiting-line systems. "Arrivals" occur in numerous ways; as machine breakdowns, as jobs coming into a jobshop, as units being assembled on a line, as orders to a warehouse, as data packets to a computer system, as calls to a call center, and so on.

Service times could be constant or probabilistic. If service times are completely random, the exponential distribution is often used for simulation purposes; however, there are several other possibilities. It could happen that the service times are constant, but some random variability causes fluctuations in either a positive or a negative way. For example, the time it takes for a lathe to traverse a 10-centimeter shaft should always be the same. However, the material could have slight differences in hardness or the tool might wear; either event could cause different processing times. In these cases, the normal distribution might describe the service time.

A special case occurs when the phenomenon of interest seems to follow the normal probability distribution, but the random variable is restricted to be greater than or less than a certain value. In this case, the truncated normal distribution can be utilized.

The gamma and Weibull distributions are also used to model interarrival and service times. (Actually, the exponential distribution is a special case of both the gamma and the Weibull distributions.) The differences between the exponential, gamma, and Weibull distributions involve the location of the modes of the pdf's and the shapes of their tails for large and small times. The exponential distribution has its mode at the origin, but the gamma and Weibull distributions have their modes at some point (≥0) that is a function of the parameter values selected. The tail of the gamma distribution is long, like an exponential distribution; the tail of the Weibull distribution can decline more rapidly or less rapidly than that of an exponential distribution.

In practice, this means that, if there are more large service times than an exponential distribution can account for, a Weibull distribution might provide a better model of these service times.

2. *Inventory and supply-chain systems.* In realistic inventory and supply-chain systems, there are at least three random variables: (1) the number of units demanded per order or per time period, (2) the time between demands, and (3) the lead time. (The lead time is defined as the time between the placing of an order for stocking the inventory system and the receipt of that order.) In very simple mathematical models of inventory systems, demand is a constant over time, and lead time is zero, or a constant. However, in most real-world cases, and, hence, in simulation models, demand occurs randomly in time, and the number of units demanded each time a demand occurs is also random, as illustrated by Figure 5.6.

Distributional assumptions for demand and lead time in inventory theory texts are usually based on mathematical tractability, but those assumptions could be invalid in a realistic context. In practice, the lead-time distribution can often be fitted fairly well by a gamma distribution [Hadley and Whitin, 1963]. Unlike analytic models, simulation models can accommodate whatever assumptions appear most reasonable.

The geometric, Poisson, and negative binomial distributions provide a range of distribution shapes that satisfy a variety of demand patterns. The geometric distribution, which is a special case of the negative binomial, has its mode at unity, given that at least one demand has occurred. If demand data are characterized by a long tail, the negative binomial distribution might be appropriate. The Poisson distribution is often used to model demand because it is simple, it is extensively tabulated, and it is well known. The tail of the Poisson distribution is generally shorter than that of the negative binomial, which means that fewer large demands will occur if a Poisson model is used than if a negative binomial distribution is used (assuming that both models have the same mean demand).

3. *Reliability and maintainability.* Time to failure has been modeled with numerous distributions, including the exponential, gamma, and Weibull. If only random failures occur, the time-to-failure distribution may be modeled as exponential. The gamma distribution arises from modeling standby redundancy, where each component has an exponential time to failure. The Weibull distribution has been extensively used to represent time to failure, and its nature is such that it can be made to approximate many observed phenomena [Hines and Montgomery, 1990]. When there are a number of components in a system and failure is due to the most serious of a large number of defects, or possible defects, the Weibull distribution seems to do particularly well as a model. In situations where most failures are due to wear, the normal distribution might very well be appropriate [Hines and Montgomery, 1990]. The lognormal distribution has been found to be applicable in describing time to failure for some types of components.

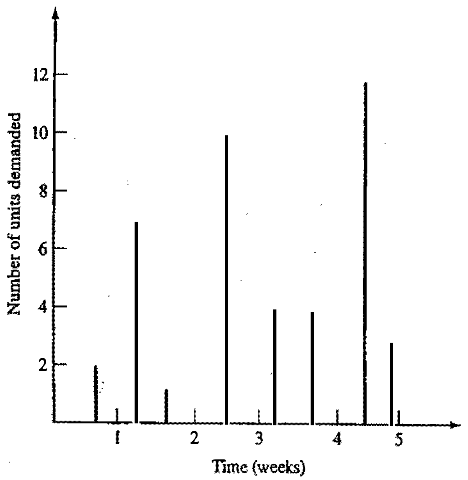

Figure 5.6 Random demands in time.

4. *Limited data.* In many instances, simulations begin before data collection has been completed. There are three distributions that have application to incomplete or limited data. These are the uniform, triangular, and beta distributions. The uniform distribution can be used when an interarrival or service time is known to be random, but no information is immediately available about the distribution [Gordon, 1975]. However, there are those who do not favor using the uniform distribution, calling it the "distribution of maximum ignorance" because it is not necessary to specify more than the continuous interval in which the random variable may occur. The triangular distribution can be used when assumptions are made about the minimum, maximum, and modal values of the random variable. Finally, the beta distribution provides a variety of distributional forms on the unit interval, ones that, with appropriate modification, can be shifted to any desired interval. The uniform distribution is a special case of the beta distribution. Pegden, Shannon, and Sadowski [1995] discuss the subject of limited data in some detail, and we include further discussion in Chapter 9.

5. *Other distributions.* Several other distributions may be useful in discrete-system simulation. The Bernoulli and binomial distributions are two discrete distributions which might describe phenomena of interest. The hyperexponential distribution is similar to the exponential distribution, but its greater variability might make it useful in certain instances.

5.3 DISCRETE DISTRIBUTIONS

Discrete random variables are used to describe random phenomena in which only integer values can occur. Numerous examples were given in Section 5.2—for example, demands for inventory items. Four distributions are described in the following subsections.

1. *Bernoulli trials and the Bernoulli distribution.* Consider an experiment consisting of n trials, each of which can be a success or a failure. Let $X_j = 1$ if the jth experiment resulted in a success, and let $X_j = 0$ if the jth experiment resulted in a failure. The n Bernoulli trials are called a Bernoulli process if the trials are independent, each trial has only two possible outcomes (success or failure), and the probability of a success remains constant from trial to trial. Thus,

$$p(x_1, x_2, \ldots, x_n) = p_1(x_1) \cdot p_2(x_2) \cdots p_n(x_n)$$

and

$$p_j(x_j) = p(x_j) = \begin{cases} p, & x_j = 1, j = 1, 2, \ldots, n \\ 1 - p = q, & x_j = 0, j = 1, 2, \ldots, n \\ 0, & \text{otherwise} \end{cases} \tag{5.11}$$

For one trial, the distribution given in Equation (5.11) is called the Bernoulli distribution. The mean and variance of X_j are calculated as follows:

$$E(X_j) = 0 \cdot q + 1 \cdot p = p$$

and

$$V(X_j) = [(0^2 \cdot q) + (1^2 \cdot p)] - p^2 = p(1 - p)$$

2. *Binomial distribution.* The random variable X that denotes the number of successes in n Bernoulli trials has a binomial distribution given by $p(x)$, where

$$p(x)=\begin{cases}\binom{n}{x}p^x q^{n-x}, & x=0,1,2,\ldots,n\\ 0, & \text{otherwise}\end{cases} \tag{5.12}$$

Equation (5.12) is motivated by computing the probability of a particular outcome with all the successes, each denoted by S, occurring in the first x trials, followed by the $n-x$ failures, each denoted by an F—that is,

$$P(\overbrace{SSS\ldots\ldots\ldots SS}^{x \text{ of these}}\overbrace{FF\ldots\ldots\ldots FF}^{n-x \text{ of these}})=p^x q^{n-x}$$

where $q=1-p$. There are

$$\binom{n}{x}=\frac{n!}{x!(n-x)!}$$

outcomes having the required number of S's and F's. Therefore, Equation (5.12) results. An easy approach to calculating the mean and variance of the binomial distribution is to consider X as a sum of n independent Bernoulli random variables, each with mean p and variance $p(1-p)=pq$. Then,

$$X=X_1+X_2+\cdots+X_n$$

and the mean, $E(X)$, is given by

$$E(X)=p+p+\cdots+p=np \tag{5.13}$$

and the variance $V(X)$ is given by

$$V(X)=pq+pq+\cdots+pq=npq \tag{5.14}$$

Example 5.10

A production process manufactures computer chips on the average at 2% nonconforming. Every day, a random sample of size 50 is taken from the process. If the sample contains more than two nonconforming chips, the process will be stopped. Compute the probability that the process is stopped by the sampling scheme.

Consider the sampling process as $n=50$ Bernoulli trials, each with $p=0.02$; then the total number of nonconforming chips in the sample, X, would have a binomial distribution given by

$$p(x)=\begin{cases}\binom{50}{x}(0.02)^x(0.98)^{50-x}, & x=0,1,2,\ldots,50\\ 0, & \text{otherwise}\end{cases}$$

It is much easier to compute the right-hand side of the following identity to compute the probability that more than two nonconforming chips are found in a sample:

$$P(X>2)=1-P(X\le 2)$$

The probability $P(X\le 2)$ is calculated from

$$\begin{aligned}P(X\le 2)&=\sum_{x=0}^{2}\binom{50}{x}(0.02)^x(0.98)^{50-x}\\&=(0.98)^{50}+50(0.02)(0.98)^{49}+1225(0.02)^2(0.98)^{48}\\&=0.92\end{aligned}$$

Thus, the probability that the production process is stopped on any day, based on the sampling process, is approximately 0.08. The mean number of nonconforming chips in a random sample of size 50 is given by

$$E(X)=np=50(0.02)=1$$

and the variance is given by

$$V(X)=npq=50(0.02)(0.98)=0.98$$

The cdf for the binomial distribution has been tabulated by Banks and Heikes [1984] and others. The tables decrease the effort considerably for computing probabilities such as $P(a<X\le b)$. Under certain conditions on n and p, both the Poisson distribution and the normal distribution may be used to approximate the binomial distribution [Hines and Montgomery, 1990].

3. *Geometric and Negative Binomial distributions.* The geometric distribution is related to a sequence of Bernoulli trials; the random variable of interest, X, is defined to be the number of trials to achieve the first success. The distribution of X is given by

$$p(x)=\begin{cases}q^{x-1}p, & x=1,2,\ldots\\ 0, & \text{otherwise}\end{cases} \tag{5.15}$$

The event $\{X=x\}$ occurs when there are $x-1$ failures followed by a success. Each of the failures has an associated probability of $q=1-p$, and each success has probability p. Thus,

$$P(FFF\cdots FS)=q^{x-1}p$$

The mean and variance are given by

$$E(X)=\frac{1}{p} \tag{5.16}$$

and

$$V(X)=\frac{q}{p^2} \tag{5.17}$$

More generally, the negative binomial distribution is the distribution of the number of trials until the kth success, for $k=1,2,\ldots$. If Y has a negative binomial distribution with parameters p and k, then the distribution of Y is given by

$$p(y)=\begin{cases}\binom{y-1}{k-1}q^{y-k}p^k, & y=k,k+1,k+2,\ldots\\ 0, & \text{otherwise}\end{cases} \tag{5.18}$$

Because we can think of the negative binomial random variable Y as the sum of k independent geometric random variables, it is easy to see that $E(Y)=k/p$ and $V(X)=kq/p^2$.

Example 5.11

Forty percent of the assembled ink-jet printers are rejected at the inspection station. Find the probability that the first acceptable ink-jet printer is the third one inspected. Considering each inspection as a Bernoulli trial with $q = 0.4$ and $p = 0.6$ yields

$$p(3) = 0.4^2(0.6) = 0.096$$

Thus, in only about 10% of the cases is the first acceptable printer the third one from any arbitrary starting point. To determine the probability that the third printer inspected is the second acceptable printer, we use the negative binomial distribution (5.18),

$$p(3) = \binom{3-1}{2-1} 0.4^{3-2}(0.6)^2 = \binom{2}{1} 0.4(0.6)^2 = 0.288$$

4. *Poisson distribution.* The Poisson distribution describes many random processes quite well and is mathematically quite simple. The Poisson distribution was introduced in 1837 by S. D. Poisson in a book concerning criminal and civil justice matters. (The title of this rather old text is "Recherches sur la probabilite des jugements en matiere criminelle et en matiere civile." Evidently, the rumor handed down through generations of probability theory professors concerning the origin of the Poisson distribution is just not true. Rumor has it that the Poisson distribution was first used to model deaths from the kicks of horses in the Prussian Army.)

The Poisson probability mass function is given by

$$p(x) = \begin{cases} \dfrac{e^{-\alpha}\alpha^x}{x!}, & x = 0, 1, \ldots \\ 0, & \text{otherwise} \end{cases} \tag{5.19}$$

where $\alpha > 0$. One of the important properties of the Poisson distribution is that the mean and variance are both equal to α, that is,

$$E(X) = \alpha = V(X)$$

The cumulative distribution function is given by

$$F(x) = \sum_{i=0}^{x} \frac{e^{-\alpha}\alpha^i}{i!} \tag{5.20}$$

The pmf and cdf for a Poisson distribution with $\alpha = 2$ are shown in Figure 5.7. A tabulation of the cdf is given in Table A.4.

Example 5.12

A computer repair person is "beeped" each time there is a call for service. The number of beeps per hour is known to occur in accordance with a Poisson distribution with a mean of $\alpha = 2$ per hour. The probability of three beeps in the next hour is given by Equation (5.19) with $x = 3$, as follows:

$$p(3) = \frac{e^{-2}2^3}{3!} = \frac{(0.135)(8)}{6} = 0.18$$

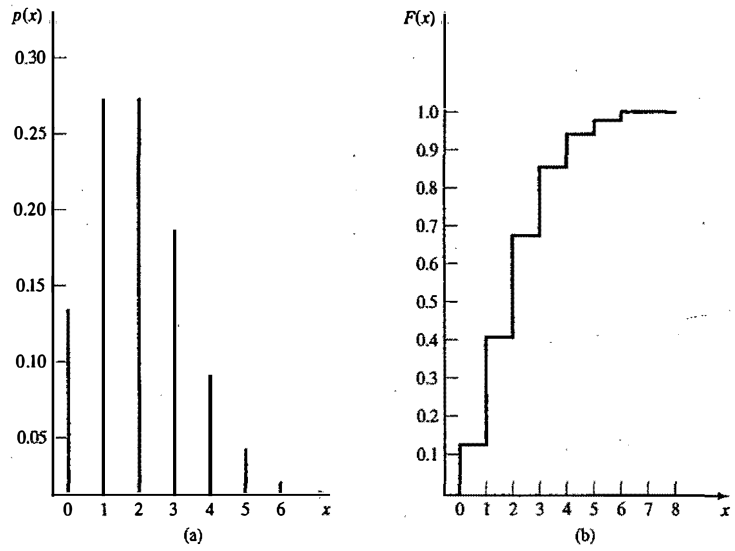

Figure 5.7 Poisson pmf and cdf.

This same result can be read from the left side of Figure 5.7 or from Table A.4 by computing

$$F(3) - F(2) = 0.857 - 0.677 = 0.18$$

Example 5.13

In Example 5.12, find the probability of two or more beeps in a 1-hour period.

$$\begin{aligned} P(2 \text{ or more}) &= 1 - p(0) - p(1) = 1 - F(1) \\ &= 1 - 0.406 = 0.594 \end{aligned}$$

The cumulative probability, $F(1)$, can be read from the right side of Figure 5.7 or from Table A.4.

Example 5.14

The lead-time demand in an inventory system is the accumulation of demand for an item from the point at which an order is placed until the order is received—that is,

$$L = \sum_{i=1}^{T} D_i \tag{5.21}$$

where L is the lead-time demand, D_i, is the demand during the ith time period, and T is the number of time periods during the lead time. Both D_i and T may be random variables.

An inventory manager desires that the probability of a stockout not exceed a certain fraction during the lead time. For example, it may be stated that the probability of a shortage during the lead time not exceed 5%.

If the lead-time demand is Poisson distributed, the determination of the reorder point is greatly facilitated. The reorder point is the level of inventory at which a new order is placed.

Assume that the lead-time demand is Poisson distributed with a mean of $\alpha = 10$ units and that 95% protection from a stockout is desired. Thus, it is desired to find the smallest value of x such that the probability that the lead-time demand does not exceed x is greater than or equal to 0.95. Using Equation (5.20) requires finding the smallest x such that

$$F(x) = \sum_{i=0}^{x} \frac{e^{-10}10^i}{i!} \geq 0.95$$

The desired result occurs at $x = 15$, which can be found by using Table A.4 or by computation of $p(0), p(1), \ldots$.

5.4 CONTINUOUS DISTRIBUTIONS

Continuous random variables can be used to describe random phenomena in which the variable of interest can take on any value in some interval—for example, the time to failure or the length of a rod. Eight distributions are described in the following subsections.

1. *Uniform distribution.* A random variable X is uniformly distributed on the interval (a, b) if its pdf is given by

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{otherwise} \end{cases} \tag{5.22}$$

The cdf is given by

$$F(x) = \begin{cases} 0, & x < a \\ \dfrac{x-a}{b-a}, & a \leq x < b \\ 1, & x \geq b \end{cases} \tag{5.23}$$

Note that

$$P(x_1 < X < x_2) = F(x_2) - F(x_1) = \frac{x_2 - x_1}{b-a}$$

is proportional to the length of the interval, for all x_1 and x_2 satisfying $a \leq x_1 < x_2 \leq b$. The mean and variance of the distribution are given by

$$E(X) = \frac{a+b}{2} \tag{5.24}$$

and

$$V(X) = \frac{(b-a)^2}{12} \tag{5.25}$$

The pdf and cdf when $a = 1$ and $b = 6$ are shown in Figure 5.8.

The uniform distribution plays a vital role in simulation. Random numbers, uniformly distributed between zero and 1, provide the means to generate random events. Numerous methods for generating uniformly distributed random numbers have been devised; some will be discussed in Chapter 7. Uniformly distributed

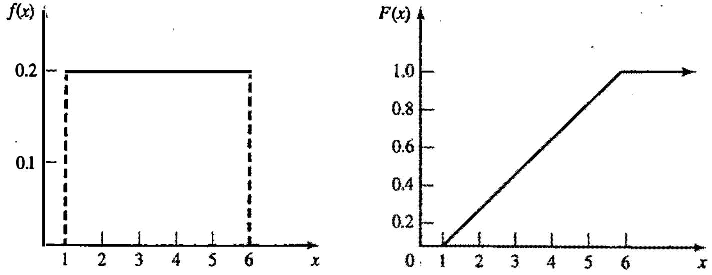

Figure 5.8 pdf and cdf for uniform distribution.

random numbers are then used to generate samples of random variates from all other distributions, as will be discussed in Chapter 8.

Example 5.15

A simulation of a warehouse operation is being developed. About every 3 minutes, a call comes for a forklift truck operator to proceed to a certain location. An initial assumption is made that the time between calls (arrivals) is uniformly distributed with a mean of 3 minutes. By Equation (5.25), the uniform distribution with a mean of 3 and the greatest possible variability would have parameter values of $a = 0$ and $b = 6$ minutes. With very limited data (such as a mean of approximately 3 minutes) plus the knowledge that the quantity of interest is variable in a random fashion, the uniform distribution with greatest variance can be assumed, at least until more data are available.

Example 5.16

A bus arrives every 20 minutes at a specified stop beginning at 6:40 A.M. and continuing until 8:40 A.M. A certain passenger does not know the schedule, but arrives randomly (uniformly distributed) between 7:00 A.M. and 7:30 A.M. every morning. What is the probability that the passenger waits more than 5 minutes for a bus?

The passenger has to wait more than 5 minutes only if the arrival time is between 7:00 A.M. and 7:15 A.M. or between 7:20 A.M. and 7:30 A.M. If X is a random variable that denotes the number of minutes past 7:00 A.M. that the passenger arrives, the desired probability is

$$P(0 < X < 15) + P(20 < X < 30)$$

Now, X is a uniform random variable on (0,30). Therefore, the desired probability is given by

$$F(15) + F(30) - F(20) = \frac{15}{30} + 1 - \frac{20}{30} = \frac{5}{6}$$

2. *Exponential distribution.* A random variable X is said to be exponentially distributed with parameter $\lambda > 0$ if its pdf is given by

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & \text{elsewhere} \end{cases} \tag{5.26}$$

The density function is shown in Figures 5.9 and 5.3. Figure 5.9 also shows the cdf.

The exponential distribution has been used to model interarrival times when arrivals are completely random and to model service times that are highly variable. In these instances, λ is a rate: arrivals per hour

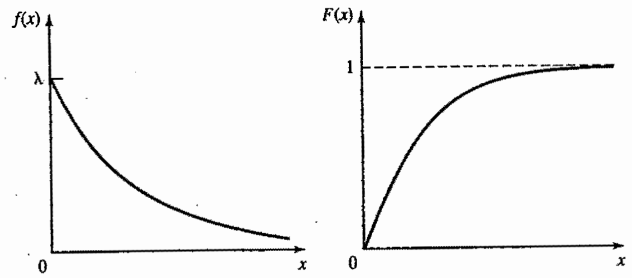

Figure 5.9 Exponential density function and cumulative distribution function.

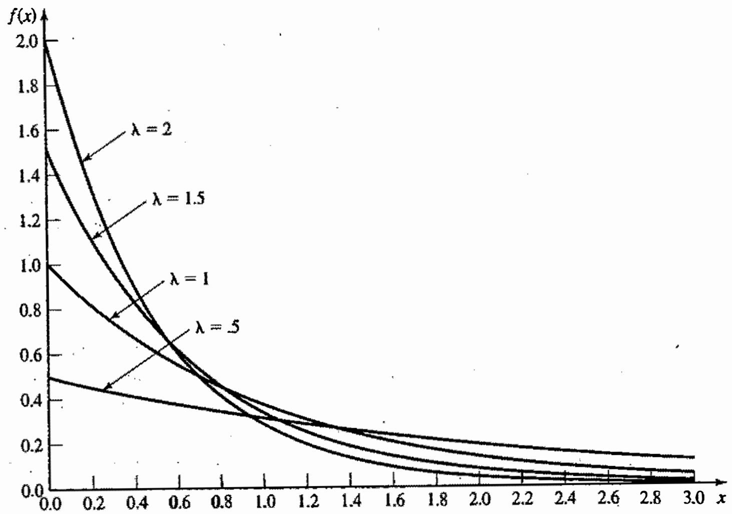

Figure 5.10 pdfs for several exponential distributions.

or services per minute. The exponential distribution has also been used to model the lifetime of a component that fails catastrophically (instantaneously), such as a light bulb; then λ is the failure rate.

Several different exponential pdf's are shown in Figure 5.10. The value of the intercept on the vertical axis is always equal to the value of λ. Note also that all pdf's eventually intersect. (Why?)

The exponential distribution has mean and variance given by

$$E(X)=\frac{1}{\lambda} \quad \text{and} \quad V(X)=\frac{1}{\lambda^2} \tag{5.27}$$

Thus, the mean and standard deviation are equal. The cdf can be exhibited by integrating Equation (5.26) to obtain

$$F(x)=\begin{cases} 0, & x<0 \\ \int_0^x \lambda e^{-\lambda t}\,dt = 1-e^{-\lambda x}, & x \geq 0 \end{cases} \tag{5.28}$$

Example 5.17

Suppose that the life of an industrial lamp, in thousands of hours, is exponentially distributed with failure rate $\lambda = 1/3$ (one failure every 3000 hours, on the average). The probability that the lamp will last longer than its mean life, 3000 hours, is given by $P(X > 3) = 1 - P(X \leq 3) = 1 - F(3)$. Equation (5.28) is used to compute $F(3)$, obtaining

$$P(X>3)=1-(1-e^{-3/3})=e^{-1}=0.368$$

Regardless of the value of λ, this result will always be the same! That is, the probability that an exponential random variable is greater than its mean is 0.368, for any value of λ.

The probability that the industrial lamp will last between 2000 and 3000 hours is computed as

$$P(2 \leq X \leq 3) = F(3) - F(2)$$

Again, from the cdf given by Equation (5.28),

$$\begin{aligned} F(3)-F(2) &= (1-e^{-3/3})-(1-e^{-2/3}) \\ &= -0.368+0.513=0.145 \end{aligned}$$

One of the most important properties of the exponential distribution is that it is "memoryless," which means that, for all $s \geq 0$ and $t \geq 0$,

$$P(X > s+t | X > s) = P(X > t) \tag{5.29}$$

Let X represent the life of a component (a battery, light bulb, computer chip, laser, etc.) and assume that X is exponentially distributed. Equation (5.29) states that the probability that the component lives for at least $s + t$ hours, given that it has survived s hours, is the same as the initial probability that it lives for at least t hours. If the component is alive at time s (if $X > s$), then the distribution of the remaining amount of time that it survives, namely $X - s$, is the same as the original distribution of a new component. That is, the component does not "remember" that it has already been in use for a time s. A used component is as good as new.

That Equation (5.29) holds is shown by examining the conditional probability

$$P(X>s+t \mid X>s)=\frac{P(X>s+t)}{P(X>s)} \tag{5.30}$$

Equation (5.28) can be used to determine the numerator and denominator of Equation (5.30), yielding

$$\begin{aligned} P(X>s+t \mid X>s) &= \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} \\ &= P(X>t) \end{aligned}$$

Example 5.18

Find the probability that the industrial lamp in Example 5.17 will last for another 1000 hours, given that it is operating after 2500 hours. This determination can be found using Equations (5.29) and (5.28), as follows:

$$P(X>3.5 \mid X>2.5) = P(X>1) = e^{-1/3} = 0.717$$

Example 5.18 illustrates the *memoryless* property—namely, that a used component that follows an exponential distribution is as good as a new component. The probability that a new component will have a life greater than 1000 hours is also equal to 0.717. Stated in general, suppose that a component which has a lifetime that follows the exponential distribution with parameter λ is observed and found to be operating at an arbitrary time. Then, the distribution of the remaining lifetime is also exponential with parameter λ. The exponential distribution is the only continuous distribution that has the memoryless property. (The geometric distribution is the only discrete distribution that possesses the memoryless property.)

3. *Gamma distribution.* A function used in defining the gamma distribution is the gamma function, which is defined for all $\beta > 0$ as

$$\Gamma(\beta) = \int_0^\infty x^{\beta-1} e^{-x} dx \tag{5.31}$$

By integrating Equation (5.31) by parts, it can be shown that

$$\Gamma(\beta) = (\beta - 1)\Gamma(\beta - 1) \tag{5.32}$$

If β is an integer, then, by using $\Gamma(1) = 1$ and applying Equation (5.32), it can be seen that

$$\Gamma(\beta) = (\beta - 1)! \tag{5.33}$$

The gamma function can be thought of as a generalization of the factorial notion to all positive numbers, not just integers.

A random variable X is gamma distributed with parameters β and θ if its pdf is given by

$$f(x) = \begin{cases} \dfrac{\beta\theta}{\Gamma(\beta)}(\beta\theta x)^{\beta-1} e^{-\beta\theta x}, & x > 0 \\ 0, & \text{otherwise} \end{cases} \tag{5.34}$$

β is called the shape parameter, and θ is called the scale parameter. Several gamma distributions for $\theta = 1$ and various values of β are shown in Figure 5.10a.

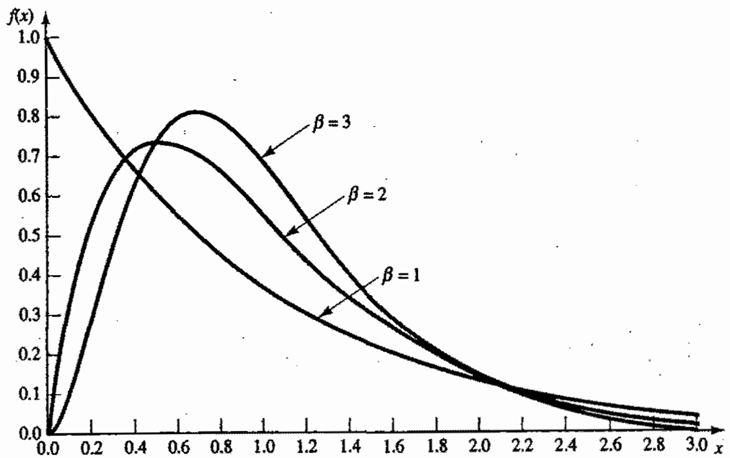

Figure 5.10a

The mean and variance of the gamma distribution are given by

$$E(X) = \frac{1}{\theta} \tag{5.35}$$

and

$$V(X) = \frac{1}{\beta\theta^2} \tag{5.36}$$

The cdf of X is given by

$$F(x) = \begin{cases} 1 - \displaystyle\int_x^\infty \frac{\beta\theta}{\Gamma(\beta)}(\beta\theta t)^{\beta-1} e^{-\beta\theta t} dt, & x > 0 \\ 0, & x \le 0 \end{cases} \tag{5.37}$$

When β is an integer, the gamma distribution is related to the exponential distribution in the following manner: If the random variable, X, is the sum of β independent, exponentially distributed random variables, each with parameter $\beta\theta$, then X has a gamma distribution with parameters β and θ. Thus, if

$$X = X_1 + X_2 + \cdots + X_\beta \tag{5.38}$$

where the pdf of X_j is given by

$$g(x_j) = \begin{cases} (\beta\theta)e^{-\beta\theta x_j}, & x \ge 0 \\ 0, & \text{otherwise} \end{cases}$$

and the X_j are mutually independent, then X has the pdf given in Equation (5.34). Note that, when $\beta = 1$, an exponential distribution results. This result follows from Equation (5.38) or from letting $\beta = 1$ in Equation (5.34).

4. *Erlang distribution.* The pdf given by Equation (5.34) is often referred to as the Erlang distribution of order (or number of phases) k when $\beta = k$, an integer. Erlang was a Danish telephone engineer who was an early developer of queueing theory. The Erlang distribution could arise in the following context: Consider a series of k stations that must be passed through in order to complete the servicing of a customer. An additional customer cannot enter the first station until the customer in process has negotiated all the stations. Each station has an exponential distribution of service time with parameter $k\theta$. Equations (5.35) and (5.36), which state the mean and variance of a gamma distribution, are valid regardless of the value of β. However, when $\beta = k$, an integer, Equation (5.38) may be used to derive the mean of the distribution in a fairly straightforward manner. The expected value of the sum of random variables is the sum of the expected value of each random variable. Thus,

$$E(X) = E(X_1) + E(X_2) + \cdots + E(X_k)$$

The expected value of each of the exponentially distributed X_j is given by $1/k\theta$. Thus,

$$E(X) = \frac{1}{k\theta} + \frac{1}{k\theta} + \cdots + \frac{1}{k\theta} = \frac{1}{\theta}$$

If the random variables X_j are independent, the variance of their sum is the sum of the variances, or

$$V(X) = \frac{1}{(k\theta)^2} + \frac{1}{(k\theta)^2} + \cdots + \frac{1}{(k\theta)^2} = \frac{1}{k\theta^2}$$

When $\beta = k$, a positive integer, the cdf given by Equation (5.37) may be integrated by parts, giving

$$F(x) = \begin{cases} 1 - \sum_{i=0}^{k-1} \frac{e^{-k\theta x}(k\theta x)^i}{i!}, & x > 0 \\ 0, & x \le 0 \end{cases} \tag{5.39}$$

which is the sum of Poisson terms with mean $\alpha = k\theta x$. Tables of the cumulative Poisson distribution may be used to evaluate the cdf when the shape parameter is an integer.

Example 5.19

A college professor of electrical engineering is leaving home for the summer, but would like to have a light burning at all times to discourage burglars. The professor rigs up a device that will hold two light bulbs. The device will switch the current to the second bulb if the first bulb fails. The box in which the light bulbs are packaged says, "Average life 1000 hours, exponentially distributed." The professor will be gone 90 days (2160 hours). What is the probability that a light will be burning when the summer is over and the professor returns?

The probability that the system will operate at least x hours is called the reliability function $R(x)$:

$$R(x) = 1 - F(x)$$

In this case, the total system lifetime is given by Equation (5.38) with $\beta = k = 2$ bulbs and $k\theta = 1/1000$ per hour, so $\theta = 1/2000$ per hour. Thus, $F(2160)$ can be determined from Equation (5.39) as follows:

$$F(2160) = 1 - \sum_{i=0}^{1} \frac{e^{-(2)(1/2000)(2160)}[(2)(1/2000)(2160)]^i}{i!}$$
$$= 1 - e^{-2.16} \sum_{i=0}^{1} \frac{(2.16)^i}{i!} = 0.636$$

Therefore, the chances are about 36% that a light will be burning when the professor returns.

Example 5.20

A medical examination is given in three stages by a physician. Each stage is exponentially distributed with a mean service time of 20 minutes. Find the probability that the exam will take 50 minutes or less. Also, compute the expected length of the exam. In this case, $k = 3$ stages and $k\theta = 1/20$, so that $\theta = 1/60$ per minute. Thus, $F(50)$ can be calculated from Equation (5.39) as follows:

$$F(50) = 1 - \sum_{i=0}^{2} \frac{e^{-(3)(1/60)(50)}[(3)(1/60)(50)]^i}{i!}$$
$$= 1 - \sum_{i=0}^{2} \frac{e^{-5/2}(5/2)^i}{i!}$$

The cumulative Poisson distribution, shown in Table A.4, can be used to calculate that

$$F(50) = 1 - 0.543 = 0.457$$

The probability is 0.457 that the exam will take 50 minutes or less. The expected length of the exam is found from Equation (5.35):

$$E(X) = \frac{1}{\theta} = \frac{1}{1/60} = 60 \text{ minutes}$$

In addition, the variance of X is $V(X) = 1/\beta\theta^2 = 1200$ minutes^2—incidentally, the mode of the Erlang distribution is given by

$$\text{Mode} = \frac{k-1}{k\theta} \tag{5.40}$$

Thus, the modal value in this example is

$$\text{Mode} = \frac{3-1}{3(1/60)} = 40 \text{ minutes}$$

5. *Normal distribution.* A random variable X with mean $-\infty < \mu < \infty$ and variance $\sigma^2 > 0$ has a normal distribution if it has the pdf

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right], \quad -\infty < x < \infty \tag{5.41}$$

The normal distribution is used so often that the notation $X \sim N(\mu, \sigma^2)$ has been adopted by many authors to indicate that the random variable X is normally distributed with mean μ and variance σ^2. The normal pdf is shown in Figure 5.11.

Some of the special properties of the normal distribution are listed here:

1. $\lim_{x\to-\infty} f(x) = 0$ and $\lim_{x\to\infty} f(x) = 0$; the value of $f(x)$ approaches zero as x approaches negative infinity and, similarly, as x approaches positive infinity.
2. $f(\mu - x) = f(\mu + x)$; the pdf is symmetric about μ.
3. The maximum value of the pdf occurs at $x = \mu$; the mean and mode are equal.

The cdf for the normal distribution is given by

$$F(x) = P(X \le x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right] dt \tag{5.42}$$

It is not possible to evaluate Equation (5.42) in closed form. Numerical methods could be used, but it appears that it would be necessary to evaluate the integral for each pair (μ, σ^2). However, a transformation of

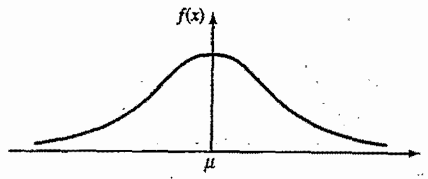

Figure 5.11 pdf of the normal distribution.

variables, $z = (t - \mu)/\sigma$, allows the evaluation to be independent of μ and σ. If $X \sim N(\mu, \sigma^2)$, let $Z = (X - \mu)/\sigma$ to obtain

$$F(x) = P(X \le x) = P\left(Z \le \frac{x-\mu}{\sigma}\right)$$

$$= \int_{-\infty}^{(x-\mu)/\sigma} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$$

$$= \int_{-\infty}^{(x-\mu)/\sigma} \phi(z)dz = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

The pdf

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad -\infty < z < \infty \tag{5.44}$$

is the pdf of a normal distribution with mean zero and variance 1. Thus, $Z \sim N(0, 1)$ and it is said that Z has a standard normal distribution. The standard normal distribution is shown in Figure 5.12. The cdf for the standard normal distribution is given by

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt \tag{5.45}$$

Equation (5.45) has been widely tabulated. The probabilities $\Phi(z)$ for $Z \ge 0$ are given in Table A.3. Several examples are now given that indicate how Equation (5.43) and Table A.3 are used.

Example 5.21

Suppose that it is known that $X \sim N(50, 9)$. Compute $F(56) = P(X \le 56)$. Using Equation (5.43) get

$$F(56) = \Phi\left(\frac{56-50}{3}\right) = \Phi(2) = 0.9772$$

from Table A.3. The intuitive interpretation is shown in Figure 5.13. Figure 5.13(a) shows the pdf of $X \sim N(50, 9)$ with the specific value, $x_0 = 56$, marked. The shaded portion is the desired probability. Figure 5.13(b) shows the standard normal distribution or $Z \sim N(0, 1)$ with the value 2 marked; $x_0 = 56$ is 2σ (where $\sigma = 3$) greater than the mean. It is helpful to make both sketches such as those in Figure 5.13 to avoid confusion in figuring out required probabilities.

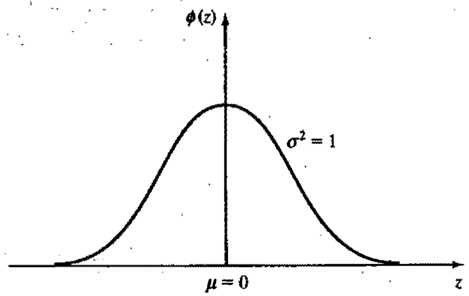

Figure 5.12 pdf of the standard normal distribution.

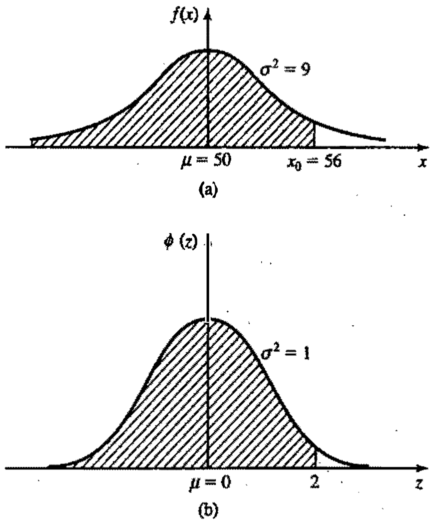

Figure 5.13 Transforming to the standard normal distribution.

Example 5.22

The time in hours required to load an oceangoing vessel, X, is distributed as $N(12,4)$. The probability that the vessel will be loaded in less than 10 hours is given by $F(10)$, where

$$F(10) = \Phi\left(\frac{10-12}{2}\right) = \Phi(-1) = 0.1587$$

The value of $\Phi(-1) = 0.1587$ is looked up in Table A.3 by using the symmetry property of the normal distribution. Note that $\Phi(1) = 0.8413$. The complement of 0.8413, or 0.1587, is contained in the tail, the shaded portion of the standard normal distribution shown in Figure 5.14(a). In Figure 5.14(b), the symmetry property is used to work out the shaded region to be $\Phi(-1) = 1 - \Phi(1) = 0.1587$. [From this logic, it can be seen that $\Phi(2) = 0.9772$ and $\Phi(-2) = 1 - \Phi(2) = 0.0228$. In general, $\Phi(-x) = 1 - \Phi(x)$.]

The probability that 12 or more hours will be required to load the ship can also be discovered by inspection, by using the symmetry property of the normal pdf and the mean as shown by Figure 5.15. The shaded portion of Figure 5.15(a) shows the problem as originally stated [i.e., evaluate $P(X < 12)$]. Now, $P(X > 12) = 1 - F(12)$. The standardized normal in Figure 5.15 (b) is used to evaluate $F(12) = \Phi(0) = 0.50$. Thus, $P(X > 12) = 1 - 0.50 = 0.50$. [The shaded portions in both Figure 5.15(a) and (b) contain 0.50 of the area under the normal pdf.]

The probability that between 10 and 12 hours will be required to load a ship is given by

$$P(10 \le X \le 12) = F(12) - F(10) = 0.5000 - 0.1587 = 0.3413$$

using earlier results presented in this example. The desired area is shown in the shaded portion of Figure 5.16(a). The equivalent problem shown in terms of the standardized normal distribution is shown in Figure 5.16(b). The probability statement is $F(12) - F(10) = \Phi(0) - \Phi(-1) = 0.5000 - 0.1587 - 0.3413$, from Table A.3.

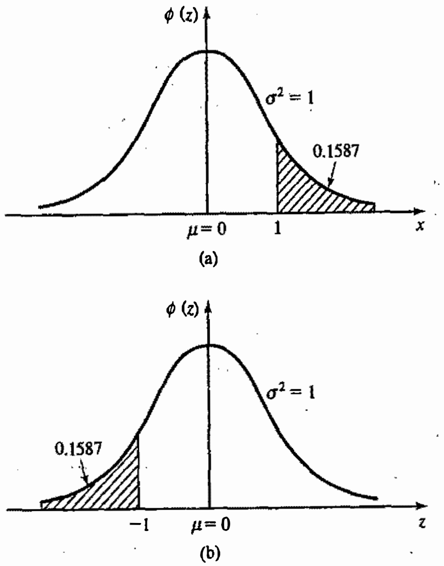

Figure 5.14 Using the symmetry property of the normal distribution.

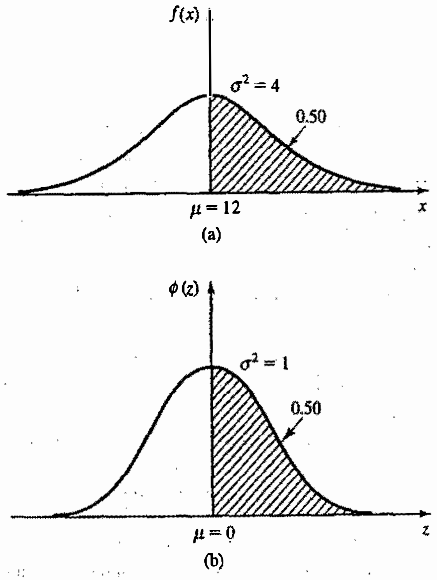

Figure 5.15 Evaluation of probability by inspection.

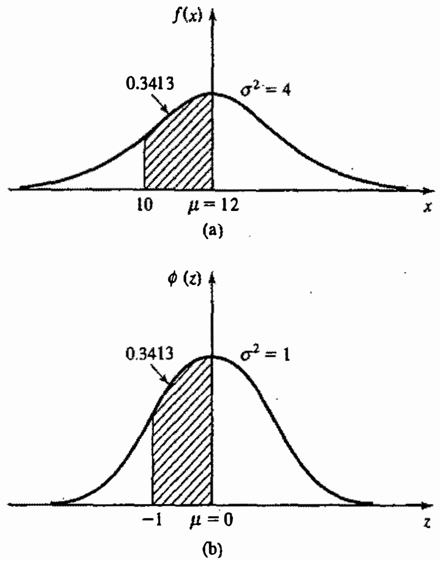

Figure 5.16 Transformation to standard normal for vessel-loading problem.

Example 5.23

The time to pass through a queue to begin self-service at a cafeteria has been found to be $N(15, 9)$. The probability that an arriving customer waits between 14 and 17 minutes is computed as follows:

$$P(14 \le X \le 17) = F(17) - F(14) = \Phi\left(\frac{17-15}{3}\right) - \Phi\left(\frac{14-15}{3}\right)$$
$$= \Phi(0.667) - \Phi(-0.333)$$

The shaded area shown in Figure 5.17(a) represents the probability $F(17) - F(14)$. The shaded area shown in Figure 5.17(b) represents the equivalent probability, $\Phi(0.667) - \Phi(-0.333)$, for the standardized normal distribution. From Table A.3, $\Phi(0.667) = 0.7476$. Now, $\Phi(-0.333) = 1 - \Phi(0.333) = 1 - 0.6304 = 0.3696$. Thus, $\Phi(0.667) - \Phi(-0.333) = 0.3780$. The probability is 0.3780 that the customer will pass through the queue in a time between 14 and 17 minutes.

Example 5.24

Lead-time demand, X, for an item is approximated by a normal distribution having mean 25 and variance 9. It is desired to compute the value for lead time that will be exceeded only 5% of the time. Thus, the problem is to find x_0 such that $P(X > x_0) = 0.05$, as shown by the shaded area in Figure 5.18(a). The equivalent problem is shown as the shaded area in Figure 5.18(b). Now,

$$P(X > x_0) = P\left(Z > \frac{x_0 - 25}{3}\right) = 1 - \Phi\left(\frac{x_0 - 25}{3}\right) = 0.05$$

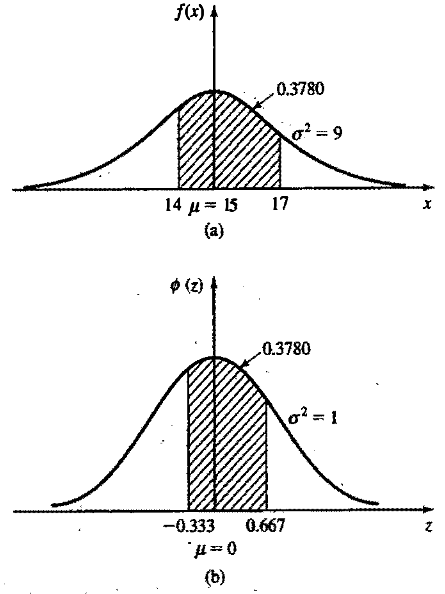

Figure 5.17 Transformation to standard normal for cafeteria problem.

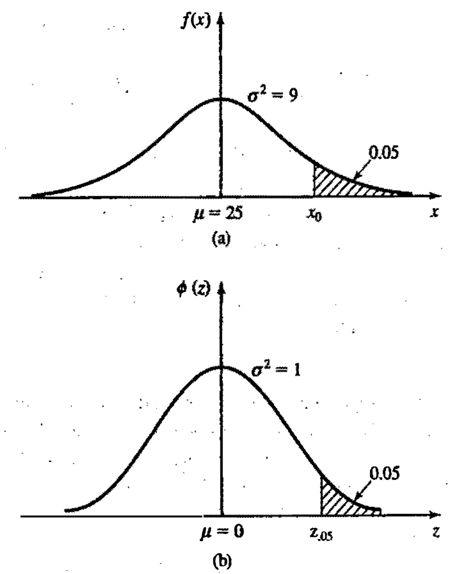

Figure 5.18 Finding x_0 for lead-time–demand problem.

or, equivalently,

$$\Phi\left(\frac{x_0-25}{3}\right)=0.95$$

From Table A.3, it can be seen that $\Phi(1.645) = 0.95$. Thus, x_0 can be found by solving

$$\frac{x_0-25}{3}=1.645$$

or

$$x_0 = 29.935$$

Therefore, in only 5% of the cases will demand during lead time exceed available inventory if an order to purchase is made when the stock level reaches 30.

6. *Weibull distribution.* The random variable X has a Weibull distribution if its pdf has the form

$$f(x)=\begin{cases}\dfrac{\beta}{\alpha}\left(\dfrac{x-\nu}{\alpha}\right)^{\beta-1}\exp\left[-\left(\dfrac{x-\nu}{\alpha}\right)^{\beta}\right], & x\geq \nu\\ 0, & \text{otherwise}\end{cases} \tag{5.46}$$

The three parameters of the Weibull distribution are ν $(-\infty < \nu < \infty)$, which is the location parameter; $\alpha\,(\alpha > 0)$, which is the scale parameter; and $\beta\,(\beta > 0)$, which is the shape parameter. When $\nu = 0$, the Weibull pdf becomes

$$f(x)=\begin{cases}\dfrac{\beta}{\alpha}\left(\dfrac{x}{\alpha}\right)^{\beta-1}\exp\left[-\left(\dfrac{x}{\alpha}\right)^{\beta}\right], & x\geq 0\\ 0, & \text{otherwise}\end{cases} \tag{5.47}$$

Figure 5.19 shows several Weibull densities when $\nu = 0$ and $\alpha = 1$. When $\beta = 1$, the Weibull distribution is reduced to

$$f(x)=\begin{cases}\dfrac{1}{\alpha}e^{-x/\alpha}, & x\geq 0\\ 0, & \text{otherwise}\end{cases}$$

which is an exponential distribution with parameter $\lambda = 1/\alpha$.

The mean and variance of the Weibull distribution are given by the following expressions:

$$E(X)=\nu+\alpha\Gamma\left(\frac{1}{\beta}+1\right) \tag{5.48}$$

$$V(X)=\alpha^2\left[\Gamma\left(\frac{2}{\beta}+1\right)-\left[\Gamma\left(\frac{1}{\beta}+1\right)\right]^2\right] \tag{5.49}$$

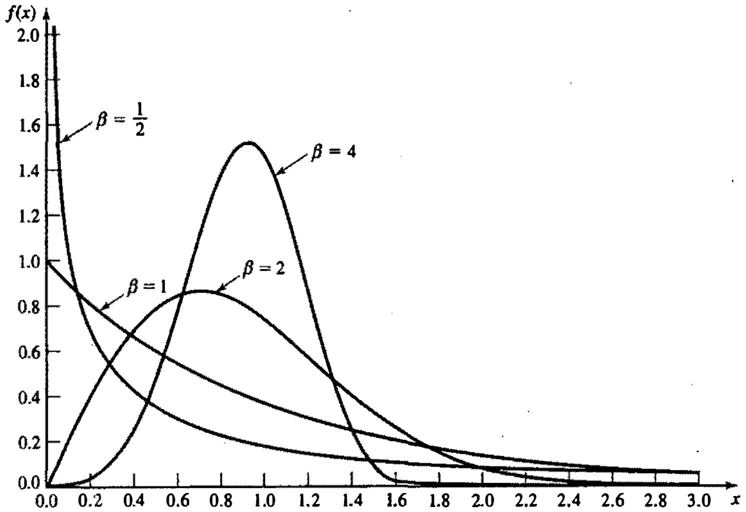

Figure 5.19 Weibull pdfs for $v=0; \alpha=\frac{1}{2}; \beta=\frac{1}{2}, 1, 2, 4$.

where $\Gamma(.)$ is defined by Equation (5.31). Thus, the location parameter, v, has no effect on the variance; however, the mean is increased or decreased by v. The cdf of the Weibull distribution is given by

$$F(x)\begin{cases} 0, & x < v \\ 1-\exp\left[-\left(\frac{x-v}{\alpha}\right)^{\beta}\right], & x \ge v \end{cases} \tag{5.50}$$

Example 5.25

The time to failure for a component screen is known to have a Weibull distribution with $v = 0$, $\beta = 1/3$, and $\alpha = 200$ hours. The mean time to failure is given by Equation (5.48) as

$$E(X) = 200\Gamma(3 + 1) = 200(3!) = 1200 \text{ hours}$$

The probability that a unit fails before 2000 hours is computed from Equation (5.50) as

$$F(2000) = 1-\exp\left[-\left(\frac{2000}{200}\right)^{1/3}\right]$$
$$= 1 - e^{-\sqrt[3]{10}} = 1 - e^{-2.15} = 0.884$$

Example 5.26

The time it takes for an aircraft to land and clear the runway at a major international airport has a Weibull distribution with $v = 1.34$ minutes, $\beta = 0.5$, and $\alpha = 0.04$ minute. Find the probability that an incoming

airplane will take more than 1.5 minutes to land and clear the runway. In this case $P(X > 1.5)$ is computed as follows:

$$P(X \le 1.5) = F(1.5)$$
$$= 1-\exp\left[-\left(\frac{1.5-1.34}{0.04}\right)^{0.5}\right]$$
$$= 1 - e^{-2} = 1 - 0.135 = 0.865$$

Therefore, the probability that an aircraft will require more than 1.5 minutes to land and clear the runway is 0.135.

7. *Triangular distribution.* A random variable X has a triangular distribution if its pdf is given by

$$f(x) = \begin{cases} \frac{2(x-a)}{(b-a)(c-a)}, & a \le x \le b \\ \frac{2(c-x)}{(c-b)(c-a)}, & b < x \le c \\ 0, & \text{elsewhere} \end{cases} \tag{5.51}$$

where $a \le b \le c$. The mode occurs at $x = b$. A triangular pdf is shown in Figure 5.20. The parameters (a, b, c) can be related to other measures, such as the mean and the mode, as follows:

$$E(X) = \frac{a+b+c}{3} \tag{5.52}$$

From Equation (5.52) the mode can be determined as

$$\text{Mode} = b = 3E(X) - (a + c) \tag{5.53}$$

Because $a \le b \le c$,

$$\frac{2a+c}{3} \le E(X) \le \frac{a+2c}{3}$$

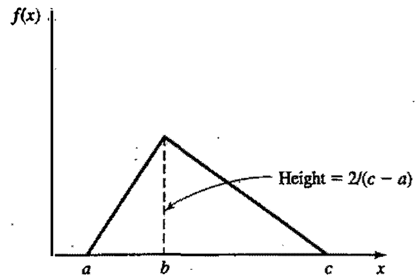

Figure 5.20 pdf of the triangular distribution.

The mode is used more often than the mean to characterize the triangular distribution. As is shown in Figure 5.20, its height is $2/(c-a)$ above the x axis. The variance, $V(X)$, of the triangular distribution is left as an exercise for the student. The cdf for the triangular distribution is given by

$$F(x)=\begin{cases}0, & x\le a\\ \dfrac{(x-a)^2}{(b-a)(c-a)}, & a<x\le b\\ 1-\dfrac{(c-x)^2}{(c-b)(c-a)}, & b<x\le c\\ 1, & x>c\end{cases} \tag{5.54}$$

Example 5.27

The central processing unit requirements, for programs that will execute, have a triangular distribution with $a = 0.05$ millisecond, $b = 1.1$ milliseconds, and $c = 6.5$ milliseconds. Find the probability that the CPU requirement for a random program is 2.5 milliseconds or less. The value of $F(2.5)$ is from the portion of the cdf in the interval (0.05, 1.1) plus that portion in the interval (1.1, 2.5). By using Equation (5.54), both portions can be addressed at one time, to yield

$$F(2.5)=1-\frac{(6.5-2.5)^2}{(6.5-0.05)(6.5-1.1)}=0.541$$

Thus, the probability is 0.541 that the CPU requirement is 2.5 milliseconds or less.

Example 5.28

An electronic sensor evaluates the quality of memory chips, rejecting those that fail. Upon demand, the sensor will give the minimum and maximum number of rejects during each hour of production over the past 24 hours. The mean is also given. Without further information, the quality control department has assumed that the number of rejected chips can be approximated by a triangular distribution. The current dump of data indicates that the minimum number of rejected chips during any hour was zero, the maximum was 10, and the mean was 4. Given that $a = 0$, $c = 10$, and $E(X) = 4$, the value of b can be found from Equation (5.53):

$$b = 3(4) - (0 + 10) = 2$$

The height of the mode is $2/(10 - 0) = 0.2$. Thus, Figure 5.21 can be drawn.

The median is the point at which 0.5 of the area is to the left and 0.5 is to the right. The median in this example is 3.7, also shown on Figure 5.21. Finding the median of the triangular distribution requires an initial location of the value to the left or to the right of the mode. The area to the left of the mode is computed from Equation (5.54) as

$$F(2)=\frac{2^2}{20}=0.2$$

Thus, the median is between b and c. Setting $F(x) = 0.5$ in Equation (5.54) and solving for x = median yields

$$0.5=1-\frac{(10-x)^2}{(10)(8)}$$

with

$$x = 3.7$$

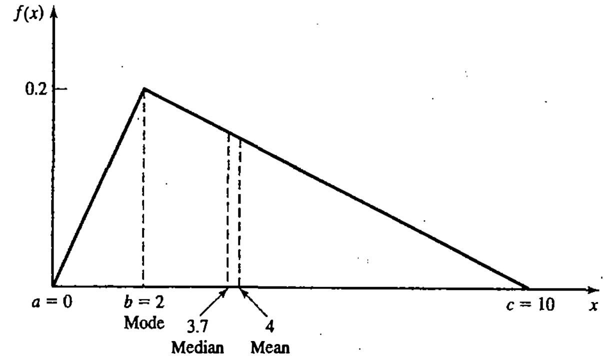

Figure 5.21 Mode, median, and mean for triangular distribution.

This example clearly shows that the mean, mode, and median are not necessarily equal.

8. *Lognormal distribution.* A random variable X has a lognormal distribution if its pdf is given by

$$f(x)=\begin{cases}\dfrac{1}{\sqrt{2\pi}\sigma x}\exp\left[-\dfrac{(\ln x-\mu)^2}{2\sigma^2}\right], & x>0\\ 0, & \text{otherwise}\end{cases} \tag{5.55}$$

where $\sigma^2 > 0$ The mean and variance of a lognormal random variable are

$$E(X)=e^{\mu+\sigma^2/2} \tag{5.56}$$

$$V(X)=e^{2\mu+\sigma^2}(e^{\sigma^2}-1) \tag{5.57}$$

Three lognormal pdf's, all having mean 1, but variances 1/2, 1, and 2, are shown in Figure 5.22.

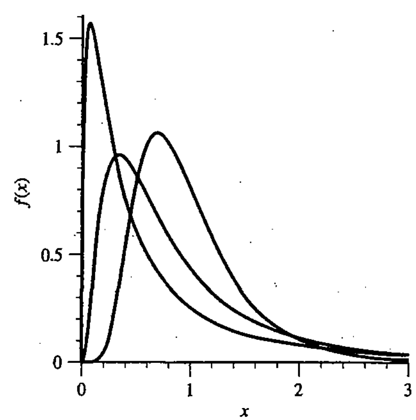

Figure 5.22 pdf of the lognormal distribution.

Notice that the parameters μ and σ^2 are not the mean and variance of the lognormal. These parameters come from the fact that when Y has a $N(\mu, \sigma^2)$ distribution then $X = e^Y$ has a lognormal distribution with parameters μ and σ^2. If the mean and variance of the lognormal are known to be μ_L and σ_L^2, respectively, then the parameters μ and σ^2 are given by

$$\mu = \ln\left(\frac{\mu_L^2}{\sqrt{\mu_L^2 + \sigma_L^2}}\right) \tag{5.58}$$

$$\sigma^2 = \ln\left(\frac{\mu_L^2 + \sigma_L^2}{\mu_L^2}\right) \tag{5.59}$$

Example 5.29

The rate of return on a volatile investment is modeled as having a lognormal distribution with mean 20% and standard deviation 5%. Compute the parameters for the lognormal distribution. From the information given, we have $\mu_L = 20$ and $\sigma_L^2 = 5^2$. Thus, from Equations (5.58) and (5.59),

$$\mu = \ln\left(\frac{20^2}{\sqrt{20^2 + 5^2}}\right) \doteq 2.9654$$

$$\sigma^2 = \ln\left(\frac{20^2 + 5^2}{20^2}\right) \doteq 0.06$$

9. *Beta distribution.* A random variable X is beta-distributed with parameters $\beta_1 > 0$ and $\beta_2 > 0$ if its pdf is given by

$$f(x) = \begin{cases} \dfrac{x^{\beta_1 - 1}(1-x)^{\beta_2 - 1}}{B(\beta_1, \beta_2)}, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases} \tag{5.60}$$

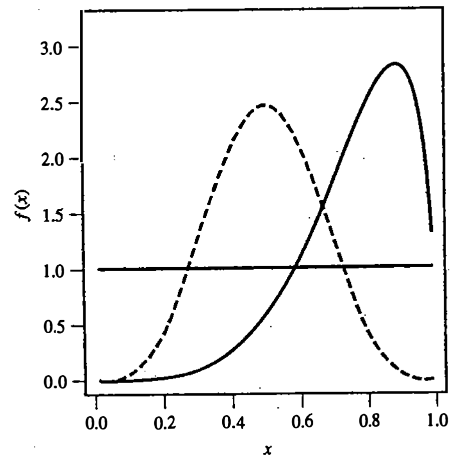

Figure 5.23 The pdf's for several beta distributions.

where $B(\beta_1, \beta_2) = \Gamma(\beta_1)\Gamma(\beta_2)/\Gamma(\beta_1 + \beta_2)$. The cdf of the beta does not have a closed form in general.

The beta distribution is very flexible and has a finite range from 0 to 1, as shown in Figure 5.23. In practice, we often need a beta distribution defined on a different range, say (a, b), with $a < b$, rather than (0, 1). This is easily accomplished by defining a new random variable

$$Y = a + (b - a)X$$

The mean and variance of Y are given by

$$a + (b-a)\left(\frac{\beta_1}{\beta_1 + \beta_2}\right) \tag{5.61}$$

and

$$(b-a)^2\left(\frac{\beta_1\beta_2}{(\beta_1 + \beta_2)^2(\beta_1 + \beta_2 + 1)}\right) \tag{5.62}$$

5.5 POISSON PROCESS

Consider random events such as the arrival of jobs at a job shop, the arrival of e-mail to a mail server, the arrival of boats to a dock, the arrival of calls to a call center, the breakdown of machines in a large factory, and so on. These events may be described by a counting function $N(t)$ defined for all $t \geq 0$. This counting function will represent the number of events that occurred in $[0, t]$. Time zero is the point at which the observation began, regardless of whether an arrival occurred at that instant. For each interval $[0, t]$, the value $N(t)$ is an observation of a random variable where the only possible values that can be assumed by $N(t)$ are the integers 0, 1, 2,....

The counting process, $\{N(t), t \geq 0\}$, is said to be a Poisson process with mean rate λ if the following assumptions are fulfilled:

1. Arrivals occur one at a time.
2. $\{N(t), t \geq 0\}$ has stationary increments: The distribution of the number of arrivals between t and $t + s$ depends only on the length of the interval s, not on the starting point t. Thus, arrivals are completely at random without rush or slack periods.
3. $\{N(t), t \geq 0\}$ has independent increments: The number of arrivals during nonoverlapping time intervals are independent random variables. Thus, a large or small number of arrivals in one time interval has no effect on the number of arrivals in subsequent time intervals. Future arrivals occur completely at random, independent of the number of arrivals in past time intervals.

If arrivals occur according to a Poisson process, meeting the three preceding assumptions, it can be shown that the probability that $N(t)$ is equal to n is given by

$$P[N(t) = n] = \frac{e^{-\lambda t}(\lambda t)^n}{n!} \quad \text{for } t \geq 0 \text{ and } n = 0, 1, 2, \ldots \tag{5.63}$$

Comparing Equation (5.63) to Equation (5.19), it can be seen that $N(t)$ has the Poisson distribution with parameter $\alpha = \lambda t$. Thus, its mean and variance are given by

$$E[N(t)] = \alpha = \lambda t = V[N(t)]$$

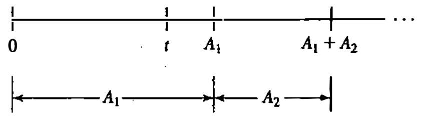

Figure 5.24 Arrival process.

For any times s and t such that $s < t$, the assumption of stationary increments implies that the random variable $N(t) - N(s)$, representing the number of arrivals in the interval from s to t, is also Poisson-distributed with mean $\lambda(t - s)$. Thus,

$$P[N(t)-N(s)=n]=\frac{e^{-\lambda(t-s)}[\lambda(t-s)]^n}{n!} \quad \text{for } n = 0, 1, 2, \ldots$$

and

$$E[N(t) - N(s)] = \lambda(t - s) = V[N(t) - N(s)]$$

Now, consider the time at which arrivals occur in a Poisson process. Let the first arrival occur at time A_1, the second occur at time $A_1 + A_2$, and so on, as shown in Figure 5.24. Thus, $A_1, A_2, \ldots$ are successive interarrival times. The first arrival occurs after time t if and only if there are no arrivals in the interval $[0, t]$, so it is seen that

$$\{A_1 > t\} = \{N(t) = 0\}$$

and, therefore,

$$P(A_1 > t) = P[N(t) = 0] = e^{-\lambda t}$$

the last equality following from Equation (5.63). Thus, the probability that the first arrival will occur in $[0, t]$ is given by

$$P(A_1 \le t) = 1 - e^{-\lambda t}$$

which is the cdf for an exponential distribution with parameter λ. Hence, A_1 is distributed exponentially with mean $E(A_1) = 1/\lambda$. It can also be shown that all interarrival times, $A_1, A_2, \ldots$, are exponentially distributed and independent with mean $1/\lambda$. As an alternative definition of a Poisson process, it can be shown that, if interarrival times are distributed exponentially and independently, then the number of arrivals by time t, say $N(t)$, meets the three previously mentioned assumptions and, therefore, is a Poisson process.

Recall that the exponential distribution is memoryless—that is, the probability of a future arrival in a time interval of length s is independent of the time of the last arrival. The probability of the arrival depends only on the length of the time interval, s. Thus, the memoryless property is related to the properties of independent and stationary increments of the Poisson process.

Additional readings concerning the Poisson process may be obtained from many sources, including Parzen [1999], Feller [1968], and Ross [2002].

Example 5.30

The jobs at a machine shop arrive according to a Poisson process with a mean of $\lambda = 2$ jobs per hour. Therefore, the interarrival times are distributed exponentially, with the expected time between arrivals being $E(A) = 1/\lambda = \frac{1}{2}$ hour.

5.5.1 Properties of a Poisson Process

Several properties of the Poisson process, discussed by Ross [2002] and others, are useful in discrete-system simulation. The first of these properties concerns random splitting. Consider a Poisson process $\{N(t), t \ge 0\}$ having rate λ, as represented by the left side of Figure 5.25.

Suppose that, each time an event occurs, it is classified as either a type I or a type II event. Suppose further that each event is classified as a type I event with probability p and type II event with probability $1-p$, independently of all other events.

Let $N_1(t)$ and $N_2(t)$ be random variables that denote, respectively, the number of type I and type II events occurring in $[0, t]$. Note that $N(t) = N_1(t) + N_2(t)$. It can be shown that $N_1(t)$ and $N_2(t)$ are both Poisson processes having rates λp and $\lambda(1 - p)$, as shown in Figure 5.25. Furthermore, it can be shown that the two processes are independent.

Example 5.31: (Random Splitting)

Suppose that jobs arrive at a shop in accordance with a Poisson process having rate λ. Suppose further that each arrival is marked "high priority" with probability 1/3 and "low priority" with probability 2/3. Then a type I event would correspond to a high-priority arrival and a type II event would correspond to a low-priority arrival. If $N_1(t)$ and $N_2(t)$ are as just defined, both variables follow the Poisson process, with rates $\lambda/3$ and $2\lambda/3$, respectively.

Example 5.32

The rate in Example 5.31 is $\lambda = 3$ per hour. The probability that no high-priority jobs will arrive in a 2-hour period is given by the Poisson distribution with parameter $\alpha = \lambda pt = 2$. Thus,

$$P(0)=\frac{e^{-2}2^0}{0!}=0.135$$

Now, consider the opposite situation from random splitting, namely the pooling of two arrival streams. The process of interest is illustrated in Figure 5.26. It can be shown that, if $N_i(t)$ are random variables representing independent Poisson processes with rates λ_i, for $i = 1$ and 2, then $N(t) = N_1(t) + N_2(t)$ is a Poisson process with rate $\lambda_1 + \lambda_2$.

Example 5.33: (Pooled Process)

A Poisson arrival stream with $\lambda_1 = 10$ arrivals per hour is combined (or pooled) with a Poisson arrival stream with $\lambda_2 = 17$ arrivals per hour. The combined process is a Poisson process with $\lambda = 27$ arrivals per hour.

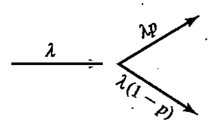

Figure 5.25 Random splitting.

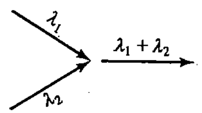

Figure 5.26 Pooled process.

5.5.2 Nonstationary Poisson Process

If we keep the Poisson Assumptions 1 and 3, but drop Assumption 2 (stationary increments) then we have a *nonstationary Poisson process* (NSPP), which is characterized by $\lambda(t)$, the arrival rate at time t. The NSPP is useful for situations in which the arrival rate varies during the period of interest, including meal times for restaurants, phone calls during business hours, and orders for pizza delivery around 6 P.M.

The key to working with a NSPP is the expected number of arrivals by time t, denoted by

$$\Lambda(t) = \int_0^t \lambda(s)ds$$

To be useful as an arrival-rate function, $\lambda(t)$ must be nonnegative and integrable. For a stationary Poisson process with rate λ we have $\Lambda(t) = \lambda t$, as expected.

Let $T_1, T_2, \ldots$ be the arrival times of stationary Poisson process $N(t)$ with $\lambda = 1$, and let $\mathcal{T}_1, \mathcal{T}_2, \ldots$ be the arrival times for a NSPP $\mathcal{N}(t)$ with arrival rate $\lambda(t)$. The fundamental relationship for working with NSPPs is the following:

$$T_i = \Lambda(\mathcal{T}_i)$$
$$\mathcal{T}_i = \Lambda^{-1}(T_i)$$

In words, an NSPP can be transformed into a stationary Poisson process with arrival rate 1, and a stationary Poisson process with arrival rate 1 can be transformed into an NSPP with rate $\lambda(t)$, and the transformation in both cases is related to $\Lambda(t)$.

Example 5.34

Suppose that arrivals to a Post Office occur at a rate of 2 per minute from 8 A.M. until 12 P.M., then drop to 1 every 2 minutes until the day ends at 4 P.M. What is the probability distribution of the number of arrivals between 11 A.M. and 2 P.M?

Let time $t = 0$ correspond to 8 A.M. Then this situation could be modeled as a NSPP $\mathcal{N}(t)$ with rate function

$$\lambda(t) = \begin{cases} 2, & 0 \le t < 4 \\ \frac{1}{2}, & 4 \le t \le 8 \end{cases}$$

The expected number of arrivals by time t is therefore

$$\Lambda(t) = \begin{cases} 2t, & 0 \le t < 4 \\ \frac{t}{2} + 6, & 4 \le t \le 8 \end{cases}$$

Notice that computing the expected number of arrivals for $4 \le t \le 8$ requires that the integration be done in two parts:

$$\Lambda(t) = \int_0^t \lambda(s)\,ds = \int_0^4 2\,ds + \int_4^t \tfrac{1}{2}\,ds = \frac{t}{2} + 6$$

Since 2 P.M. and 11 A.M. correspond to times 6 and 3, respectively, we have

$$\begin{aligned} P[\mathcal{N}(6) - \mathcal{N}(3) = k] &= P[N(\Lambda(6)) - N(\Lambda(3)) = k] \\ &= P[N(9) - N(6) = k] \\ &= \frac{e^{9-6}(9-6)^k}{k!} \\ &= \frac{e^{3}(3)^k}{k!} \end{aligned}$$

where $N(t)$ is a stationary Poisson process with arrival rate 1.

5.6 EMPIRICAL DISTRIBUTIONS

An empirical distribution, which may be either discrete or continuous in form, is a distribution whose parameters are the observed values in a sample of data. This is in contrast to parametric distribution families (such as the exponential, normal, or Poisson), which are characterized by specifying a small number of parameters such as the mean and variance. An empirical distribution may be used when it is impossible or unnecessary to establish that a random variable has any particular parametric distribution. One advantage of an empirical distribution is that nothing is assumed beyond the observed values in the sample; however, this is also a disadvantage because the sample might not cover the entire range of possible values.

Example 5.35: (Discrete)

Customers at a local restaurant arrive at lunchtime in groups of from one to eight persons. The number of persons per party in the last 300 groups has been observed; the results are summarized in Table 5.3. The relative frequencies appear in Table 5.3 and again in Figure 5.27, which provides a histogram of the data that were gathered. Figure 5.28 provides a cdf of the data. The cdf in Figure 5.28 is called the empirical distribution of the given data.

Example 5.36: (Continuous)

The time required to repair a conveyor system that has suffered a failure has been collected for the last 100 instances; the results are shown in Table 5.4. There were 21 instances in which the repair took between 0 and 0.5 hour, and so on. The empirical cdf is shown in Figure 5.29. A piecewise linear curve is formed by the connection of the points of the form $[x, F(x)]$. The points are connected by a straight line. The first connected pair is (0, 0) and (0.5, 0.21); then the points (0.5, 0.21) and (1.0, 0.33) are connected; and so on. More detail on this method is provided in Chapter 8.

Table 5.3 Arrivals per Party Distribution

Arrivals per Party	*Frequency*	*Relative Frequency*	*Cumulative Relative Frequency*
1	30	0.10	0.10
2	110	0.37	0.47
3	45	0.15	0.62
4	71	0.24	0.86
5	12	0.04	0.90
6	13	0.04	0.94
7	7	0.02	0.96
8	12	0.04	1.00

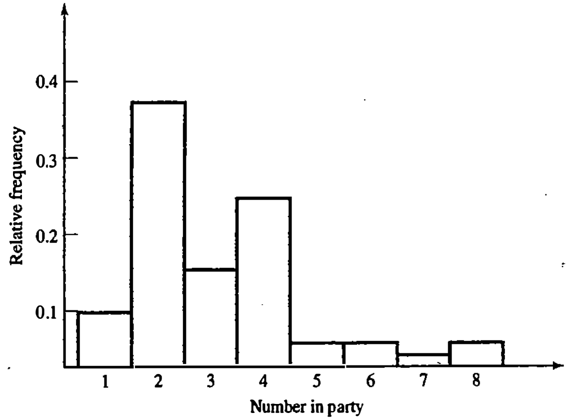

Figure 5.27 Histogram of party size.

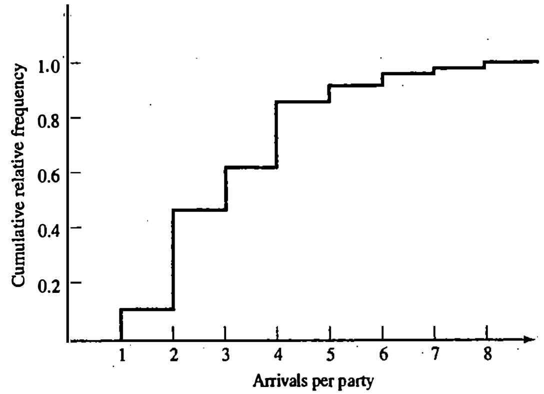

Figure 5.28 Empirical cdf of party size.

Table 5.4 Repair Times for Conveyor

Interval (Hours)	*Relative Frequency*	*Cumulative Frequency*	*Frequency*
$0 < x \le 0.5$	21	0.21	0.21
$0.5 < x \le 1.0$	12	0.12	0.33
$1.0 < x \le 1.5$	29	0.29	0.62
$1.5 < x \le 2.0$	19	0.19	0.81
$2.0 < x \le 2.5$	8	0.08	0.89
$2.5 < x \le 3.0$	11	0.11	1.00

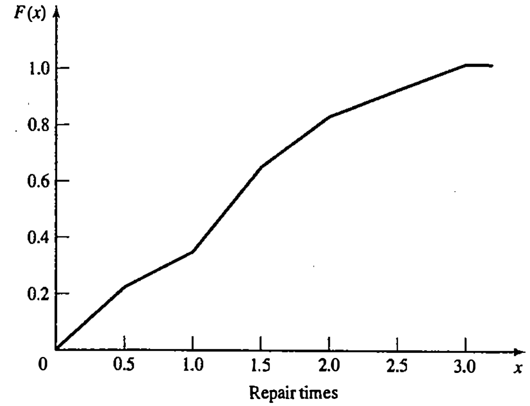

Figure 5.29 Empirical cdf for repair times.

5.7 SUMMARY

In many instances, the world the simulation analyst sees is probabilistic rather than deterministic. The purposes of this chapter were to review several important probability distributions, to familiarize the reader with the notation used in the remainder of the text, and to show applications of the probability distributions in a simulation context.

A major task in simulation is the collection and analysis of input data. One of the first steps in this task is hypothesizing a distributional form for the input data. This is accomplished by comparing the shape of the probability density function or mass function to a histogram of the data and by an understanding that certain physical processes give rise to specific distributions. (Computer software is available to assist in this effort, as will be discussed in Chapter 9.) This chapter was intended to reinforce the properties of various distributions and to give insight into how these distributions arise in practice. In addition, probabilistic models of input data are used in generating random events in a simulation.

Several features that should have made a strong impression on the reader include the differences between discrete, continuous, and empirical distributions; the Poisson process and its properties; and the versatility of the gamma and the Weibull distributions.

REFERENCES

BANKS, J., AND R. G. HEIKES [1984], *Handbook of Tables and Graphs for the Industrial Engineer and Manager*, Reston Publishing, Reston, VA.

DEVORE, J. L. [1999], *Probability and Statistics for Engineers and the Sciences*, 5th ed., Brooks/Cole, Pacific Grove, CA.

FELLER, W. [1968], *An Introduction to Probability Theory and Its Applications*, Vol. I, 3d ed., Wiley, New York.

GORDON, G. [1975], *The Application of GPSS V to Discrete System Simulation*, Prentice-Hall, Englewood Cliffs, NJ.

HADLEY, G., AND T. M. WHITIN [1963], *Analysis of Inventory Systems*, Prentice-Hall, Englewood Cliffs, NJ.

HINES, W. W., AND D. C. MONTGOMERY [1990], *Probability and Statistics in Engineering and Management Science*, 3d ed., Wiley, New York.

LAW, A. M., AND W. D. KELTON [2000], *Simulation Modeling & Analysis*, 3d ed., McGraw–Hill, New York.

PAPOULIS, A. [1990], *Probability and Statistics*, Prentice Hall, Englewood Cliffs, NJ.

PARZEN, E. [1999], *Stochastic Process*, Classics in Applied Mathematics, 24, Society for Industrial & Applied Mathematics, Philadelphia, PA.
PEGDEN, C. D., R. E. SHANNON, AND R. P. SADOWSKI [1995], *Introduction to Simulation Using SIMAN*, 2d ed., McGraw-Hill, New York.
ROSS, S. M. [2002], *Introduction to Probability Models*, 8th ed., Academic Press, New York.
WALPOLE, R. E., AND R. H. MYERS [2002], *Probability and Statistics for Engineers and Scientists*, 7th ed., Prentice Hall, Upper Saddle River, NJ.

EXERCISES

1. Of the orders a job shop receives, 25% are welding jobs and 75% are machining jobs. What is the probability that

   (a) half of the next five jobs will be machining jobs?
   (b) the next four jobs will be welding jobs?

2. Three different items are moving together in a conveyor. These items are inspected visually and defective items are removed. The previous production data are given as

	Item A	*Item B*	*Item C*
Accepted	25	280	190
Rejected	975	720	810

   What is the probability that

   (a) one item is removed at a time?
   (b) two items are removed at a time?
   (c) three items are removed simultaneously?

3. A recent survey indicated that 82% of single women aged 25 years old will be married in their lifetime. Using the binomial distribution, find the probability that two or three women in a sample of twenty will never be married.

4. The Hawks are currently winning 0.55 of their games. There are 5 games in the next two weeks. What is the probability that they will win more games than they lose?

5. Joe Coledge is the third-string quarterback for the University of Lower Alatoona. The probability that Joe gets into any game is 0.40.

   (a) What is the probability that the first game Joe enters is the fourth game of the season?
   (b) What is the probability that Joe plays in no more than two of the first five games?

6. For the random variables X_1 and X_2, which are exponentially distributed with parameter $\lambda = 1$, compute $P(X_1 + X_2 > 2)$.

7. Show that the geometric distribution is memoryless.

8. Hurricane hitting the eastern coast of India follows Poisson with a mean of 0.5 per year. Determine

   (a) the probability of more than three hurricanes hitting the Indian eastern coast in a year.
   (b) the probability of not hitting the Indian eastern coast in a year.

9. Students' arrival at a university library follows Poisson with a mean of 20 per hour. Determine

   (a) the probability that there are 50 arrivals in the next 1 hour.
   (b) the probability that no student arrives in the next 1 hour.
   (c) the probability that there are 75 arrivals in the next 2 hours.

10. Records indicate that 1.8% of the entering students at a large state university drop out of school by midterm. What is the probability that three or fewer students will drop out of a random group of 200 entering students?

11. Lane Braintwain is quite a popular student. Lane receives, on the average, four phone calls a night (Poisson distributed). What is the probability that, tomorrow night, the number of calls received will exceed the average by more than one standard deviation?

12. A car service station receives cars at the rate of 5 every hour in accordance with Poisson. What is the probability that a car will arrive 2 hours after its predecessor?

13. A random variable X that has pmf given by $p(x) = 1/(n+1)$ over the range $R_X = \{0, 1, 2, \ldots, n\}$ is said to have a discrete uniform distribution.

   (a) Find the mean and variance of this distribution. *Hint:*

$$\sum_{i=1}^{n} i = \frac{n(n+1)}{2} \text{ and } \sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}$$

   (b) If $R_X = \{a, a+1, a+2, \ldots, b\}$, compute the mean and variance of X.

14. The lifetime, in years, of a satellite placed in orbit is given by the following pdf:

$$f(x) = \begin{cases} 0.4e^{-0.4x}, & x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

   (a) What is the probability that this satellite is still "alive" after 5 years?
   (b) What is the probability that the satellite dies between 3 and 6 years from the time it is placed in orbit?

15. The cars arriving at a gas station is Poisson distributed with a mean of 10 per minute. Determine the number of pumps to be installed if the firm wants to have 50% of arriving cars as zero entries (i.e., cars serviced without waiting).

16. (The Poisson distribution can be used to approximate the binomial distribution when n is large and p is small—say, p less than 0.1. In utilizing the Poisson approximation, let $\lambda = np$.) In the production of ball bearings, bubbles or depressions occur, rendering the ball bearing unfit for sale. It has been noted that, on the average, one in every 800 of the ball bearings has one or more of these defects. What is the probability that a random sample of 4000 will yield fewer than three ball bearings with bubbles or depressions?

17. For an exponentially distributed random variable X, find the value of λ that satisfies the following relationship:

$$P(X \leq 3) = 0.9P(X \leq 4)$$

18. The time between calls to a fire service station in Chennai follows exponential with a mean of 20 hours. What is the probability that there will be no calls during the next 24 hours?

19. The time to failure of a chip follows exponential with a mean of 5000 hours.

(a) The chip is in operation for the past 1000 hours. What is the probability that the chip will be in operation for another 6000 hours?
(b) After 7000 hours of operation, what is the probability that the chip will not fail for another 2000 hours?

20. The headlight bulb of a car owned by a professor has an exponential time to failure with a mean of 100 weeks. The professor has fitted a new bulb 50 weeks ago. What is the probability that the bulb will not fuse within the next 60 weeks?

21. The service time at the college cafeteria follows exponential with a mean of 2 minutes.

(a) What is the probability that two customers in front of an arriving customer will each take less than 90 seconds to complete their transactions?
(b) What is the probability that two customers in front will finish their transactions so that an arriving customer can reach the service window within 4 minutes?

22. Determine the variance, $V(X)$, of the triangular distribution.

23. The daily demand for rice at a departmental store in thousands of kilogram is found to follow gamma distribution with shape parameter 3 and scale parameter ½. Determine the probability of demand exceeding 5000 kg on a given day.

24. When Admiral Byrd went to the North Pole, he wore battery-powered thermal underwear. The batteries failed instantaneously rather than gradually. The batteries had a life that was exponentially distributed, with a mean of 12 days. The trip took 30 days. Admiral Byrd packed three batteries. What is the probability that three batteries would be a number sufficient to keep the Admiral warm?

25. In an organization's service-complaints mail box, interarrival time of mails are exponentially distributed with a mean of 10 minutes. What is the probability that five mails will arrive in 20 minutes duration?

26. The rail shuttle cars at Atlanta airport have a dual electrical braking system. A rail car switches to the standby system automatically if the first system fails. If both systems fail, there will be a crash! Assume that the life of a single electrical braking system is exponentially distributed, with a mean of 4,000 operating hours. If the systems are inspected every 5,000 operating hours, what is the probability that a rail car will not crash before that time?

27. Suppose that cars arriving at a toll booth follow a Poisson process with a mean interarrival time of 15 seconds. What is the probability that up to one minute will elapse until three cars have arrived?

28. Suppose that an average of 30 customers per hour arrive at the Sticky Donut Shop in accordance with a Poisson process. What is the probability that more than 5 minutes will elapse before both of the next two customers walk through the door?

29. Professor Dipsy Doodle gives six problems on each exam. Each problem requires an average of 30 minutes grading time for the entire class of 15 students. The grading time for each problem is exponentially distributed, and the problems are independent of each other.

(a) What is the probability that the Professor will finish the grading in $2\frac{1}{2}$ hours or less?
(b) What is the most likely grading time?
(c) What is the expected grading time?

30. An aircraft has dual hydraulic systems. The aircraft switches to the standby system automatically if the first system fails. If both systems have failed, the plane will crash. Assume that the life of a hydraulic system is exponentially distributed, with a mean of 2000 air hours.

(a) If the hydraulic systems are inspected every 2500 hours, what is the probability that an aircraft will crash before that time?
(b) What danger would there be in moving the inspection point to 3000 hours?

31. Show that the beta distribution becomes the uniform distribution over the unit interval when $\beta_1 = \beta_2 = 1$.

32. Lead time of a product in weeks is gamma-distributed with shape parameter 2 and scale parameter 1. What is the probability that the lead time exceeds 3 weeks?

33. Lifetime of an inexpensive video card for a PC, in months, denoted by the random variable X, is gamma-distributed with $\beta = 4$ and $\theta = 1/16$. What is the probability that the card will last for at least 2 years?

34. In a statewide competitive examination for engineering admission, the register number allotted to the candidates is of the form CCNNNN, where C is a character like A, B, and C, etc., and N is a number from 0 to 9. Assume that you are scanning through the rank list (based on marks secured in the competitive examination), what is the probability that

(a) the next five entries in the list will have numbers 7000 or higher?
(b) the next three entries will have numbers greater than 3000?

35. Let X be a random variable that is normally distributed, with mean 10 and variance 4. Find the values a and b such that $P(a < X < b) = 0.90$ and $|\mu - a| = |\mu - b|$.

36. Given the following distributions,

Normal (10, 4)
Triangular (4, 10, 16)
Uniform (4, 16)
find the probability that $6 < X < 8$ for each of the distributions.

37. Demand for an item follows normal distribution with a mean of 50 units and a standard deviation of 7 units. Determine the probabilities of demand exceeding 45, 55, and 65 units.

38. The annual rainfall in Chennai is normally distributed with mean 129 cm and standard deviation 32 cm.

(a) What is the probability of getting excess rain (i.e., 140 cm and above) in a given year?
(b) What is the probability of deficient rain (i.e., 80 cm and below) in a given year?

39. Three shafts are made and assembled into a linkage. The length of each shaft, in centimeters, is distributed as follows:

Shaft 1: $N(60, 0.09)$
Shaft 2: $N(40, 0.05)$
Shaft 3: $N(50, 0.11)$

(a) What is the distribution of the length of the linkage?
(b) What is the probability that the linkage will be longer than 150.2 centimeters?

(c) The tolerance limits for the assembly are (149.83, 150.21). What proportion of assemblies are within the tolerance limits? (*Hint:* If $\{X_i\}$ are n independent normal random variables, and if X_i, has mean μ_i and variance σ_i^2, then the sum

$$Y = X_1 + X_2 + \cdots + X_n$$

is normal with mean $\sum_{i=1}^{n} \mu_i$ and variance $\sum_{i=1}^{n} \sigma_i^2$.)

40. The circumferences of battery posts in a nickel--cadmium battery are Weibull-distributed with $\nu = 3.25$ centimeters, $\beta = 1/3$, and $\alpha = 0.005$ centimeters.

(a) Find the probability that a battery post chosen at random will have a circumference larger than 3.40 centimeters.
(b) If battery posts are larger than 3.50 centimeters, they will not go through the hole provided; if they are smaller than 3.30 centimeters, the clamp will not tighten sufficiently. What proportion of posts will have to be scrapped for one of these reasons?

41. The time to failure of a nickel–cadmium battery is Weibull distributed with parameters $\nu = 0$, $\beta = 1/4$, and $\alpha = 1/2$ years.

(a) Find the fraction of batteries that are expected to fail prior to 1.5 years.
(b) What fraction of batteries are expected to last longer than the mean life?
(c) What fraction of batteries are expected to fail between 1.5 and 2.5 years?

42. The time required to assemble a component follows triangular distribution with $a = 10$ seconds and $c =$ 25 seconds. The median is 15 seconds. Compute the modal value of assembly time.

43. The time to failure (in months) of a computer follows Weibull distribution with location parameter = 0, scale parameter = 2, and shape parameter = 0.35.

(a) What is the mean time to failure?
(b) What is the probability that the computer will fail by 3 months?

44. The consumption of raw material for a fabrication firm follows triangular distribution with minimum of 200 units, maximum of 275 units, and mean of 220 units. What is the median value of raw material consumption?

45. A postal letter carrier has a route consisting of five segments with the time in minutes to complete each segment being normally distributed, with means and variances as shown:

Tennyson Place	$N(38, 16)$
Windsor Parkway	$N(99, 29)$
Knob Hill Apparments	$N(85, 25)$
Evergreen Drive	$N(73, 20)$
Chastain Shopping Center	$N(52, 12)$

In addition to the times just mentioned, the letter carrier must organize the mail at the central office, which activity requires a time that is distributed by $N(90, 25)$. The drive to the starting point of the route requires a time that is distributed $N(10, 4)$. The return from the route requires a time that is distributed $N(15, 4)$. The letter carrier then performs administrative tasks with a time that is distributed $N(30, 9)$.

(a) What is the expected length of the letter carrier's work day?
(b) Overtime occurs after eight hours of work on a given day. What is the probability that the letter carrier works overtime on any given day?

(c) What is the probability that the letter carrier works overtime on two or more days in a six-day week?
(d) What is the probability that the route will be completed within ±24 minutes of eight hours on any given day? (*Hint:* See Exercise 39.)

46. The light used in the operation theater of a hospital has two bulbs. One bulb is sufficient to get the necessary lighting. The bulbs are connected in such a way that when one fails, automatically the other gets switched on. The life of each bulb is exponentially distributed with a mean of 5000 hours and the lives of the bulbs are independent of one another. What is the probability that the combined life of the light is greater than 7000 hours?

47. High temperature in Biloxi, Mississippi on July 21, denoted by the random variable X, has the following probability density function, where X is in degrees F.

$$f(x) = \begin{cases} \dfrac{2(x-85)}{119}, & 85 \le x \le 92 \\ \dfrac{2(102-x)}{170} & 92 < x \le 102 \\ 0, & \text{otherwise} \end{cases}$$

(a) What is the variance of the temperature, $V(X)$? (If you worked Exercise 22, this is quite easy.)
(b) What is the median temperature?
(c) What is the modal temperature?

48. The time to failure of Eastinghome light bulbs is Weibull distributed with $\nu = 1.8 \times 10^3$ hours, $\beta = 1/2$, and $\alpha = 1/3 \times 10^3$ hours.

(a) What fraction of bulbs are expected to last longer than the mean lifetime?
(b) What is the median lifetime of a light bulb?

49. Let time $t = 0$ correspond to 6 A.M., and suppose that the arrival rate (in arrivals per hour) of customers to a breakfast restaurant that is open from 6 to 9 A.M. is

$$\lambda(t) = \begin{cases} 30, & 0 \le t < 1 \\ 45, & 1 \le t < 2 \\ 20, & 2 \le t \le 4 \end{cases}$$

Assuming a NSPP model is appropriate, do the following: (a) Derive $\Lambda(t)$. (b) Compute the expected number of arrivals between 6:30 and 8:30 A.M. (c) Compute the probability that there are fewer than 60 arrivals between 6:30 and 8:30 A.M.