

Selecting Input Probability Distributions

Recommended sections for a first reading: 6.1, 6.2, 6.4 through 6.7, 6.11

6.1 INTRODUCTION

To carry out a simulation using random inputs such as interarrival times or demand sizes, we have to specify their probability distributions. For example, in the simulation of the single-server queueing system in Sec. 1.4.3, the interarrival times were taken to be IID exponential random variables with a mean of 1 minute; the demand sizes in the inventory simulation of Sec. 1.5 were specified to be 1, 2, 3, or 4 items with respective probabilities $\frac{1}{6}$, $\frac{1}{3}$, $\frac{1}{3}$, and $\frac{1}{6}$. Then, given that the input random variables to a simulation model follow particular distributions, the simulation proceeds through time by generating random values from these distributions. Chapters 7 and 8 discuss methods for generating random values from various distributions and processes. Our concern in this chapter is with how the analyst might go about specifying these input probability distributions.

Almost all real-world systems contain one or more sources of randomness, as illustrated in Table 6.1. Furthermore, in Figs. 6.1 through 6.4 we show histograms for four data sets taken from actual simulation projects. Figure 6.1 corresponds to 890 machine processing times (in minutes) for an automotive manufacturer. It can be seen that the histogram has a longer right tail (positive skewness) and that the minimum value is approximately 25 minutes. In Fig. 6.2 we show a histogram for 219 interarrival times (in minutes) to a drive-up bank (see Example 6.4). Figure 6.3 displays a histogram for 856 ship-loading times (in days) (see Example 6.17). Finally, in Fig. 6.4 we give a histogram for the number of yards of paper (scaled for confidentiality reasons) on 1000 large rolls of paper used to make facial or bathroom tissue. In this case the histogram has a longer left tail (negative skewness).

TABLE 6.1
Sources of randomness for common simulation applications

Type of system	Sources of randomness
Manufacturing	Processing times, machine times to failure, machine repair times
Defense-related	Arrival times and payloads of missiles or airplanes, outcome of an engagement, miss distances for munitions
Communications	Interarrival times of messages, message types, message lengths
Transportation	Ship-loading times, interarrival times of customers to a subway

Note that none of the four histograms has a symmetric shape like that of a normal distribution, despite the fact that many simulation practitioners and simulation books widely use normal input distributions.

We saw in Sec. 4.7 that it is generally necessary to represent each source of system randomness by a probability distribution (rather than just its mean) in the simulation model. The following example shows that failure to choose the “correct” distribution can also affect the accuracy of a model’s results, sometimes drastically.

EXAMPLE 6.1. A single-server queueing system (e.g., a single machine in a factory) has exponential interarrival times with a mean of 1 minute. Suppose that 200 service times are available from the system, but their underlying probability distribution is

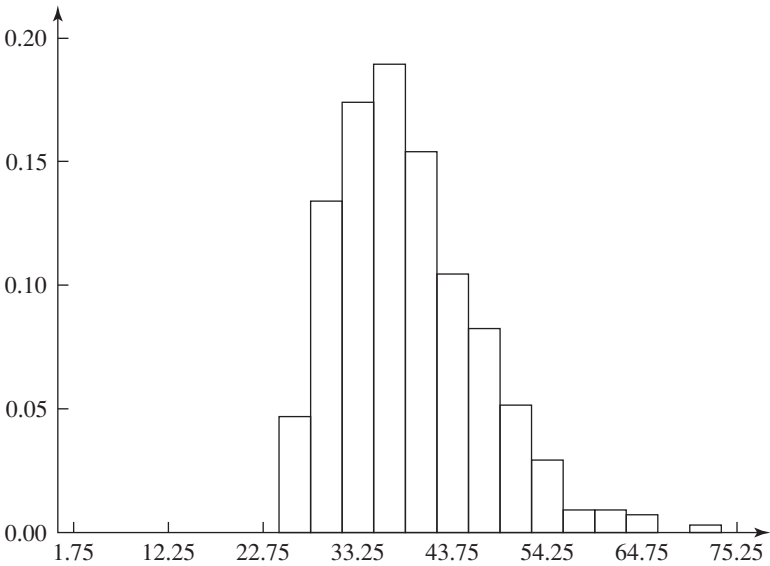


FIGURE 6.1
Histogram of 890 machine processing times for an automotive manufacturer.

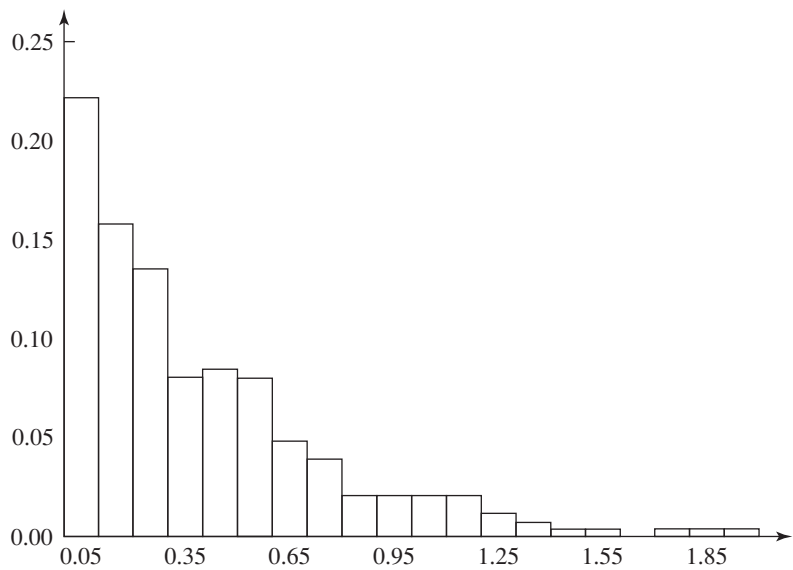


FIGURE 6.2
Histogram of 219 interarrival times to a drive-up bank.

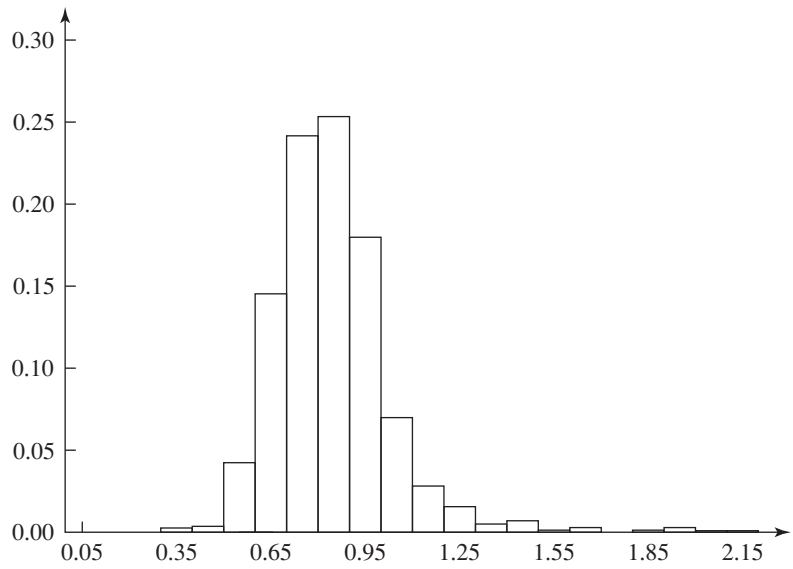


FIGURE 6.3
Histogram of 856 ship-loading times.

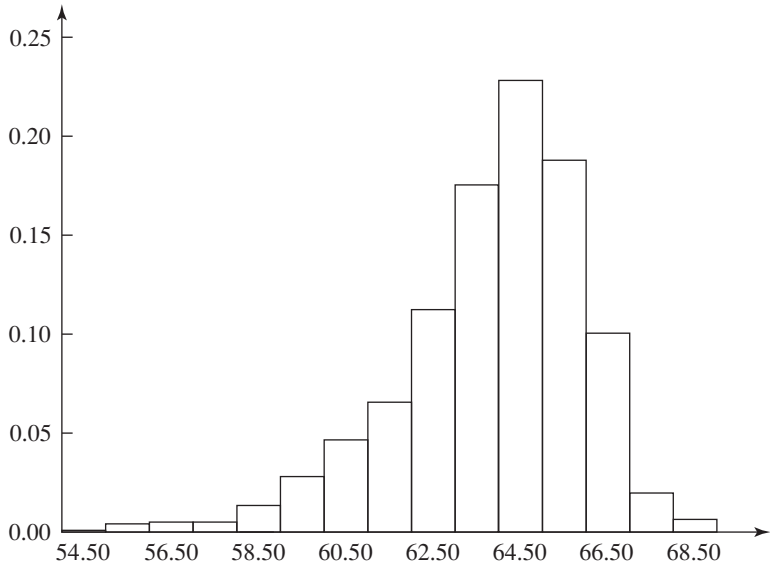


FIGURE 6.4
Histogram of the yardages for 1000 large rolls of paper for a household-products manufacturer.

unknown. Using an approach to be discussed in Sec. 6.5, we “fit” the “best” exponential, gamma, Weibull, lognormal, and normal distributions (see Sec. 6.2.2 for a discussion of these distributions) to the observed service-time data. (In the case of the exponential distribution, we chose the mean β so that the resulting distribution most closely “resembled” the available data.) We then made 100 independent simulation runs (i.e., different random numbers were used for each run, as discussed in Sec. 7.2) of the queueing system, using *each* of the five fitted distributions. (For the normal distribution, if a service time was negative, then it was generated again.) Each of the 500 simulation runs was continued until 1000 delays in queue were collected. A summary of the results from these simulation runs is given in Table 6.2. Note in the second column of the table that the average of the 100,000 delays is given for each of the service-time distributions (see Prob. 6.27). As we will see in Sec. 6.7, the Weibull distribution actually provides the best model for the service-time data. Thus, the average delay for the real system should be close to 4.36 minutes. On the other hand, the average delays for the normal and lognormal

TABLE 6.2
Simulation results for the five service-time distributions (in minutes where appropriate)

Service-time distribution	Average delay in queue	Average number in queue	Proportion of delays ≥ 20
Exponential	6.71	6.78	0.064
Gamma	4.54	4.60	0.019
Weibull	4.36	4.41	0.013
Lognormal	7.19	7.30	0.078
Normal	6.04	6.13	0.045

distributions are 6.04 and 7.19 minutes, respectively, corresponding to model output errors of 39 percent and 65 percent. This is particularly surprising for the lognormal distribution, since it has the same general shape (i.e., skewed to the right) as the Weibull distribution. However, it turns out that the lognormal distribution has a “thicker” right tail, which allows larger service times and delays to occur. The relative differences between the “tail probabilities” in column 4 of the table are even more significant. The choice of probability distributions can evidently have a large impact on the simulation output and, potentially, on the quality of the decisions made with the simulation results.

If it is possible to collect data on an input random variable of interest, these data can be used in one of the following approaches to specify a distribution (in increasing order of desirability):

1. The data values themselves are used directly in the simulation. For example, if the data represent service times, then one of the data values is used whenever a service time is needed in the simulation. This is sometimes called a *trace-driven simulation*.
2. The data values themselves are used to define an *empirical* distribution function (see Sec. 6.2.4) in some way. If these data represent service times, we would sample from this distribution when a service time is needed in the simulation.
3. Standard techniques of statistical inference are used to “fit” a *theoretical* distribution form (see Example 6.1), e.g., exponential or Poisson, to the data and to perform hypothesis tests to determine the goodness of fit. If a particular theoretical distribution with certain values for its parameters is a good model for the service-time data, then we would sample from this distribution when a service time is needed in the simulation.

Two drawbacks of approach 1 are that the simulation can only reproduce what has happened historically and that there is seldom enough data to make all the desired simulation runs. Approach 2 avoids these shortcomings since, at least for continuous data, any value between the minimum and maximum observed data points can be generated (see Sec. 8.3.16). Thus, approach 2 is generally preferable to approach 1. However, approach 1 does have its uses. For example, suppose that it is desired to compare a proposed material-handling system with the existing system for a distribution center. For each incoming order there is an arrival time, a list of the desired products, and a quantity for each product. Modeling a stream of orders for a certain period of time (e.g., for 1 month) will be difficult, if not impossible, using approach 2 or 3. Thus, in this case the existing and proposed systems will often be simulated using the historical order stream. Approach 1 is also recommended for *model validation* when model output for an existing system is compared with the corresponding output for the system itself. (See the discussion of the correlated inspection approach in Sec. 5.6.1.)

If a theoretical distribution can be found that fits the observed data reasonably well (approach 3), then this will generally be preferable to using an empirical distribution (approach 2) for the following reasons:

- An empirical distribution function may have certain “irregularities,” particularly if only a small number of data values are available. A theoretical distribution, on the other hand, “smooths out” the data and may provide information on the overall underlying distribution.

- If empirical distributions are used in the *usual* way (see Sec. 6.2.4), it is not possible to generate values outside the range of the observed data in the simulation (see Sec. 8.3.16). This is unfortunate, since many measures of performance for simulated systems depend heavily on the probability of an “extreme” event’s occurring, e.g., generation of a very large service time. With a fitted theoretical distribution, however, values outside the range of the observed data can be generated.
- There may be a compelling physical reason in some situations for using a certain theoretical distribution form as a model for a particular input random variable (see, for example, Sec. 6.12.1). Even when we are fortunate enough to have this kind of information, it is a good idea to use observed data to provide empirical support for the use of this particular distribution.
- A theoretical distribution is a compact way of representing a set of data values. Conversely, if n data values are available from a continuous distribution, then $2n$ values (e.g., data and corresponding cumulative probabilities) must be entered and stored in the computer to represent an empirical distribution in simulation packages. Thus, use of an empirical distribution will be cumbersome if the data set is large.
- A theoretical distribution is easier to change. For example, suppose that a set of interarrival times is found to be modeled well by an exponential distribution with a mean of 1 minute. If we want to determine the effect on the simulated system of increasing the arrival rate by 10 percent, then all we have to do is to change the mean of the exponential distribution to 0.909.

There are definitely situations for which no theoretical distribution will provide an adequate fit for the observed data, including the following:

- The data are a mixture of two or more heterogeneous populations (see the discussion of machine repair times in Sec. 6.4.2).
- The times to perform some task have been significantly rounded (effectively making the data discrete), and there are not enough distinct values in the sample to allow any continuous theoretical distribution to provide a good representation.

In situations where no theoretical distribution is appropriate, we recommend using an empirical distribution. Another possible drawback of theoretical distributions (e.g., lognormal) is that arbitrarily large values can be generated, albeit with a very small probability. Thus, if it is known that a random variable can never take on values larger than b , then it might be desirable to truncate the fitted theoretical distribution at b (see Sec. 6.8). For example, it might be known that a service time in a bank is extremely unlikely to exceed 15 minutes.

The remainder of this chapter discusses various topics related to the selection of input distributions. Section 6.2 discusses how theoretical distributions are parameterized, provides a compendium of relevant facts on most of the commonly used continuous and discrete distributions, and discusses how empirical distributions can be specified. In Sec. 6.3 we present techniques for determining whether the data are independent observations from some underlying distribution, which is a requirement of many of the statistical procedures in this chapter. In Secs. 6.4 through 6.6 we discuss the three basic activities in specifying a theoretical distribution on the basis

of observed data. The ExpertFit distribution-fitting software and a comprehensive example are discussed in Sec. 6.7. In Sec. 6.8 we indicate how certain of the theoretical distributions, e.g., gamma, Weibull, and lognormal, can be “shifted” away from 0 to make them better fit our observed data in some cases; we also discuss truncated distributions. We treat Bézier distributions, which are a fourth way to specify a distribution based on observed data, in Sec. 6.9. In Sec. 6.10 we describe how multivariate distributions are specified and estimated when observed data are available. In Sec. 6.11 we describe several possible methods for specifying input distributions when no data are available. Several useful probabilistic models for describing the manner in which “customers” arrive to a system are given in Sec. 6.12, while in Sec. 6.13 we present techniques for determining whether observations from different sources are homogeneous and can be merged.

The graphical plots and goodness-of-fit tests presented in this chapter were developed using the ExpertFit distribution-fitting software (see Sec. 6.7).

6.2 USEFUL PROBABILITY DISTRIBUTIONS

The purpose of this section is to discuss a variety of distributions that have been found to be useful in simulation modeling and to provide a unified listing of relevant properties of these distributions [see also Forbes et al. (2011); Johnson, Kotz, and Balakrishnan (1994, 1995); and Johnson, Kotz, and Kemp (1992)]. Section 6.2.1 provides a short discussion of common methods by which continuous distributions are defined, or parameterized. Then Secs. 6.2.2 and 6.2.3 contain compilations of several continuous and discrete distributions. Finally, Sec. 6.2.4 suggests how the data themselves can be used directly to define an empirical distribution.

6.2.1 Parameterization of Continuous Distributions

For a given family of continuous distributions, e.g., normal or gamma, there are usually several alternative ways to define, or *parameterize*, the probability density function. However, if the parameters are defined correctly, they can be classified, on the basis of their physical or geometric interpretation, as being one of three basic types: location, scale, or shape parameters.

A *location parameter* γ specifies an abscissa (x axis) location point of a distribution’s range of values; usually γ is the midpoint (e.g., the mean μ for a normal distribution) or lower endpoint (see Sec. 6.8) of the distribution’s range. (In the latter case, location parameters are sometimes called *shift parameters*.) As γ changes, the associated distribution merely shifts left or right without otherwise changing. Also, if the distribution of a random variable X has a location parameter of 0, then the distribution of the random variable $Y = X + \gamma$ has a location parameter of γ .

A *scale parameter* β determines the scale (or unit) of measurement of the values in the range of the distribution. (The standard deviation σ is a scale parameter for the normal distribution.) A change in β compresses or expands the associated

distribution without altering its basic form. Also, if the distribution of the random variable X has a scale parameter of 1, then the distribution of the random variable $Y = \beta X$ has a scale parameter of β .

A *shape parameter* α determines, distinct from location and scale, the basic form or shape of a distribution within the general family of distributions of interest. A change in α generally alters a distribution's properties (e.g., skewness) more fundamentally than a change in location or scale. Some distributions (e.g., exponential and normal) do not have a shape parameter, while others (e.g., beta) may have two.

6.2.2 Continuous Distributions

Table 6.3 gives information relevant to simulation modeling applications for 13 continuous distributions. Possible applications are given first to indicate some (certainly not all) uses of the distribution [see Hahn and Shapiro (1994) and Lawless (2003) for other applications]. Then the density function and distribution function (if it exists in simple closed form) are listed. Next is a short description of the parameters, including their possible values. The range indicates the interval where the associated random variable can take on values. Also listed are the mean (expected value), variance, and mode, i.e., the value at which the density function is maximized. MLE refers to the maximum-likelihood estimator(s) of the parameter(s), treated later in Sec. 6.5. General comments include relationships of the distribution under study to other distributions. Graphs are given of the density functions for each distribution. The notation following the name of each distribution is our abbreviation for that distribution, which includes the parameters. The symbol $\sim$ is read “is distributed as.”

Note that we have included the less familiar Johnson S_B , Johnson S_U , log-logistic, Pearson type V, and Pearson type VI distributions, because we have found that these distributions often provide a better fit to data sets than standard distributions such as gamma, lognormal, and Weibull.

TABLE 6.3
Continuous distributions

Uniform	$U(a, b)$
Possible applications	Used as a “first” model for a quantity that is felt to be randomly varying between a and b but about which little else is known. The $U(0, 1)$ distribution is essential in generating random values from all other distributions (see Chaps. 7 and 8).
Density (See Fig. 6.5)	$f(x) = \begin{cases} \frac{1}{b - a} & \text{if } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$
Distribution	$F(x) = \begin{cases} 0 & \text{if } x < a \\ \frac{x - a}{b - a} & \text{if } a \leq x \leq b \\ 1 & \text{if } b < x \end{cases}$
Parameters	a and b real numbers with $a < b$; a is a location parameter, $b - a$ is a scale parameter

TABLE 6.3 (continued)

Uniform	$U(a, b)$
Range	$[a, b]$
Mean	$\frac{a + b}{2}$
Variance	$\frac{(b - a)^2}{12}$
Mode	Does not uniquely exist
MLE	$\hat{a} = \min_{1 \leq i \leq n} X_i, \hat{b} = \max_{1 \leq i \leq n} X_i$

Comments

1. The $U(0, 1)$ distribution is a special case of the beta distribution (when $\alpha_1 = \alpha_2 = 1$).
2. If $X \sim U(0, 1)$ and $[x, x + \Delta x]$ is a subinterval of $[0, 1]$ with $\Delta x \geq 0$,

$$P(X \in [x, x + \Delta x]) = \int_x^{x+\Delta x} 1 dy = (x + \Delta x) - x = \Delta x$$

which justifies the name “uniform.”

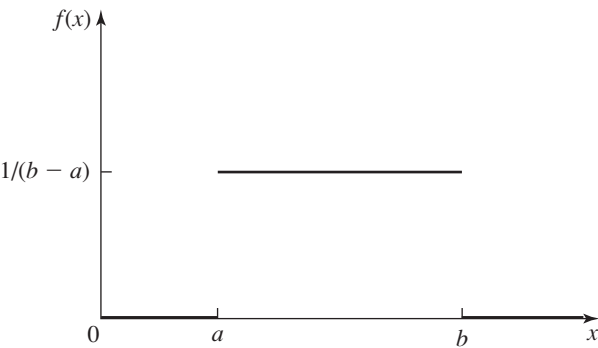


FIGURE 6.5
 $U(a, b)$ density function.

Exponential	$\text{expo}(\beta)$
Possible applications	Interarrival times of “customers” to a system that occur at a constant rate, time to failure of a piece of equipment.
Density (see Fig. 6.6)	$f(x) = \begin{cases} \frac{1}{\beta} e^{-x/\beta} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$
Distribution	$F(x) = \begin{cases} 1 - e^{-x/\beta} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$
Parameter	Scale parameter $\beta > 0$
Range	$[0, \infty)$
Mean	β
Variance	β^2
Mode	0

(continued)

TABLE 6.3 (continued)

Exponential	expo(β)
MLE	$\hat{\beta} = \bar{X}(n)$
Comments	1. The expo(β) distribution is a special case of both the gamma and Weibull distributions (for shape parameter $\alpha = 1$ and scale parameter β in both cases). 2. If $X_1, X_2, \dots, X_m$ are independent expo(β) random variables, then $X_1 + X_2 + \dots + X_m \sim \text{gamma}(m, \beta)$, also called the <i>m-Erlang(β) distribution</i>. 3. The exponential distribution is the only continuous distribution with the memoryless property (see Prob. 4.30).

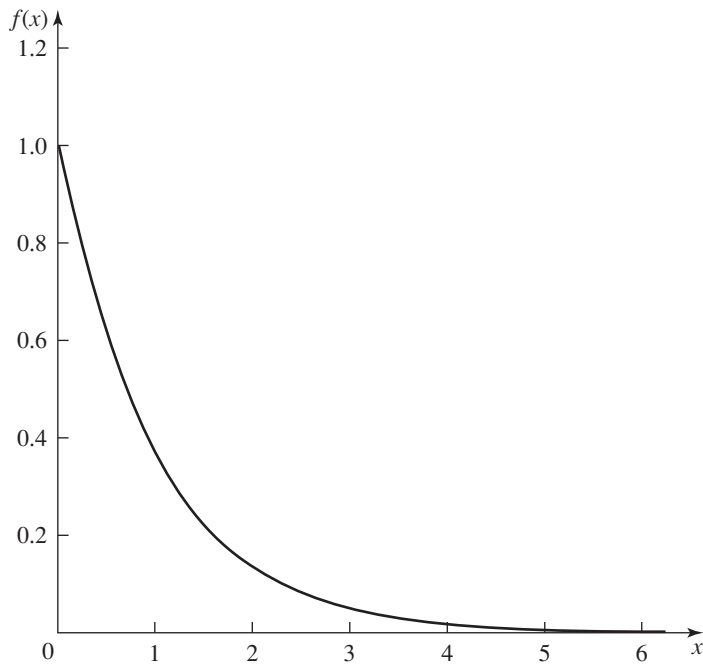


FIGURE 6.6
expo(1) density function.

Gamma	gamma(α, β)
Possible applications	Time to complete some task, e.g., customer service or machine repair
Density (see Fig. 6.7)	$f(x) = \begin{cases} \frac{\beta^{-\alpha} x^{\alpha-1} e^{-x/\beta}}{\Gamma(\alpha)} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$ where $\Gamma(\alpha)$ is the <i>gamma function</i>, defined by $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$ for any real number $z > 0$. Some properties of the gamma function: $\Gamma(z+1) = z\Gamma(z)$ for any $z > 0$, $\Gamma(k+1) = k!$ for any nonnegative integer k, $\Gamma(k+\frac{1}{2}) = \sqrt{\pi} \cdot 1 \cdot 3 \cdot 5 \cdots (2k-1)/2^k$ for any positive integer k, $\Gamma(1/2) = \sqrt{\pi}$

TABLE 6.3 (continued)

Gamma	$\text{gamma}(\alpha, \beta)$
Distribution	If α is not an integer, there is no closed form. If α is a positive integer, then $F(x) = \begin{cases} 1 - e^{-x/\beta} \sum_{j=0}^{\alpha-1} \frac{(x/\beta)^j}{j!} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$
Parameters	Shape parameter $\alpha > 0$, scale parameter $\beta > 0$
Range	$[0, \infty)$
Mean	$\alpha\beta$
Variance	$\alpha\beta^2$
Mode	$\beta(\alpha - 1)$ if $\alpha \geq 1$, 0 if $\alpha < 1$
MLE	The following two equations must be satisfied:

$$\ln \hat{\beta} + \Psi(\hat{\alpha}) = \frac{\sum_{i=1}^n \ln X_i}{n}, \quad \hat{\alpha}\hat{\beta} = \bar{X}(n)$$

which could be solved numerically. [$\Psi(\hat{\alpha}) = \Gamma'(\hat{\alpha})/\Gamma(\hat{\alpha})$ and is called the *digamma function*; Γ' denotes the derivative of Γ .] Alternatively, approximations to $\hat{\alpha}$ and $\hat{\beta}$ can be obtained by letting $T = [\ln \bar{X}(n) - \sum_{i=1}^n \ln X_i/n]^{-1}$, using Table 6.21 (see App. 6A) to obtain $\hat{\alpha}$ as a function of T , and letting $\hat{\beta} = \bar{X}(n)/\hat{\alpha}$. [See Choi and Wette (1969) for the derivation of this procedure and of Table 6.21.]

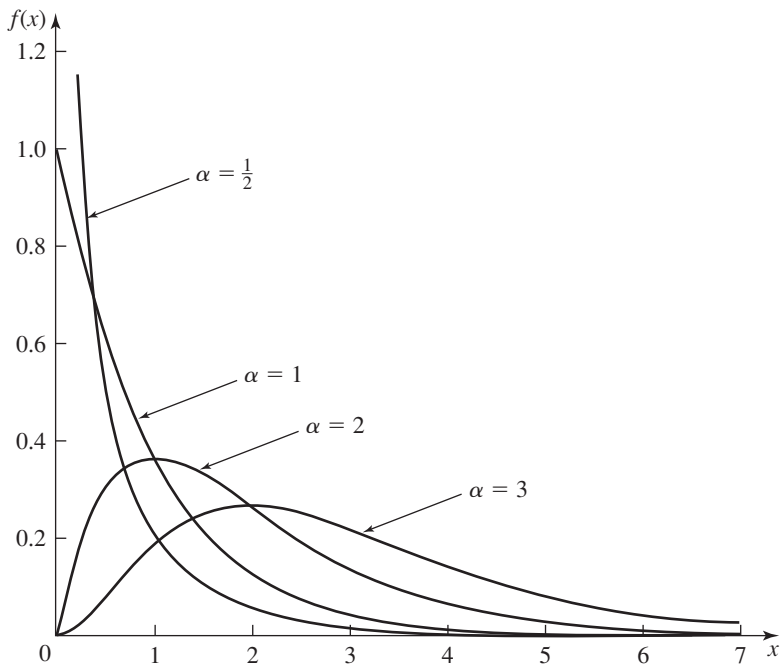


FIGURE 6.7
 $\text{gamma}(\alpha, 1)$ density functions.

(continued)

TABLE 6.3 (continued)

Gamma	gamma(α, β)
Comments	1. The expo(β) and gamma(1, β) distributions are the same. 2. For a positive integer m, the gamma(m, β) distribution is called the m-Erlang(β) distribution. 3. The chi-square distribution with k df is the same as the gamma($k/2, 2$) distribution. 4. If $X_1, X_2, \dots, X_m$ are independent random variables with $X_i \sim \text{gamma}(\alpha_i, \beta)$, then $X_1 + X_2 + \dots + X_m \sim \text{gamma}(\alpha_1 + \alpha_2 + \dots + \alpha_m, \beta)$. 5. If X_1 and X_2 are independent random variables with $X_i \sim \text{gamma}(\alpha_i, \beta)$, then $X_1/(X_1 + X_2) \sim \text{beta}(\alpha_1, \alpha_2)$. 6. $X \sim \text{gamma}(\alpha, \beta)$ if and only if $Y = 1/X$ has a Pearson type V distribution with shape and scale parameters α and $1/\beta$, denoted PT5($\alpha, 1/\beta$). 7. $\lim_{x \rightarrow 0} f(x) = \begin{cases} \infty & \text{if } \alpha < 1 \\ \frac{1}{\beta} & \text{if } \alpha = 1 \\ 0 & \text{if } \alpha > 1 \end{cases}$

Weibull	Weibull(α, β)
Possible applications	Time to complete some task, time to failure of a piece of equipment; used as a rough model in the absence of data (see Sec. 6.11)
Density (see Fig. 6.8)	$f(x) = \begin{cases} \alpha \beta^{-\alpha} x^{\alpha-1} e^{-(x/\beta)^\alpha} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$
Distribution	$F(x) = \begin{cases} 1 - e^{-(x/\beta)^\alpha} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$
Parameters	Shape parameter $\alpha > 0$, scale parameter $\beta > 0$
Range	$[0, \infty)$
Mean	$\frac{\beta}{\alpha} \Gamma\left(\frac{1}{\alpha}\right)$
Variance	$\frac{\beta^2}{\alpha} \left\{ 2\Gamma\left(\frac{2}{\alpha}\right) - \frac{1}{\alpha} \left[\Gamma\left(\frac{1}{\alpha}\right) \right]^2 \right\}$
Mode	$\begin{cases} \beta \left(\frac{\alpha-1}{\alpha} \right)^{1/\alpha} & \text{if } \alpha \geq 1 \\ 0 & \text{if } \alpha < 1 \end{cases}$
MLE	The following two equations must be satisfied:

$$\frac{\sum_{i=1}^n X_i^{\hat{\alpha}} \ln X_i}{\sum_{i=1}^n X_i^{\hat{\alpha}}} - \frac{1}{\hat{\alpha}} = \frac{\sum_{i=1}^n \ln X_i}{n}, \quad \hat{\beta} = \left(\frac{\sum_{i=1}^n X_i^{\hat{\alpha}}}{n} \right)^{1/\hat{\alpha}}$$

The first can be solved for $\hat{\alpha}$ numerically by Newton's method, and the second equation then gives $\hat{\beta}$ directly. The general recursive step for the Newton iterations is

$$\hat{\alpha}_{k+1} = \hat{\alpha}_k + \frac{A + 1/\hat{\alpha}_k - C_k/B_k}{1/\hat{\alpha}_k^2 + (B_k H_k - C_k^2)/B_k^2}$$

TABLE 6.3 (continued)

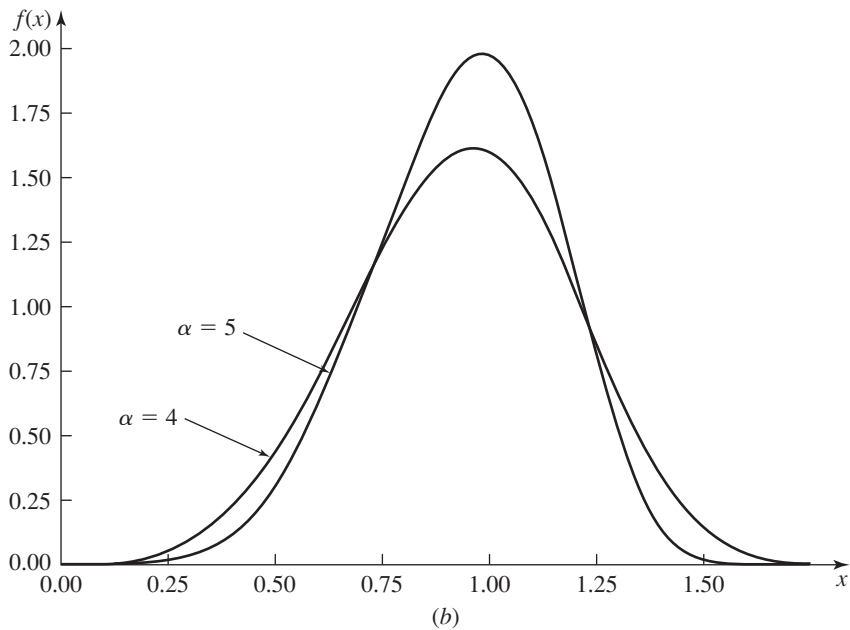
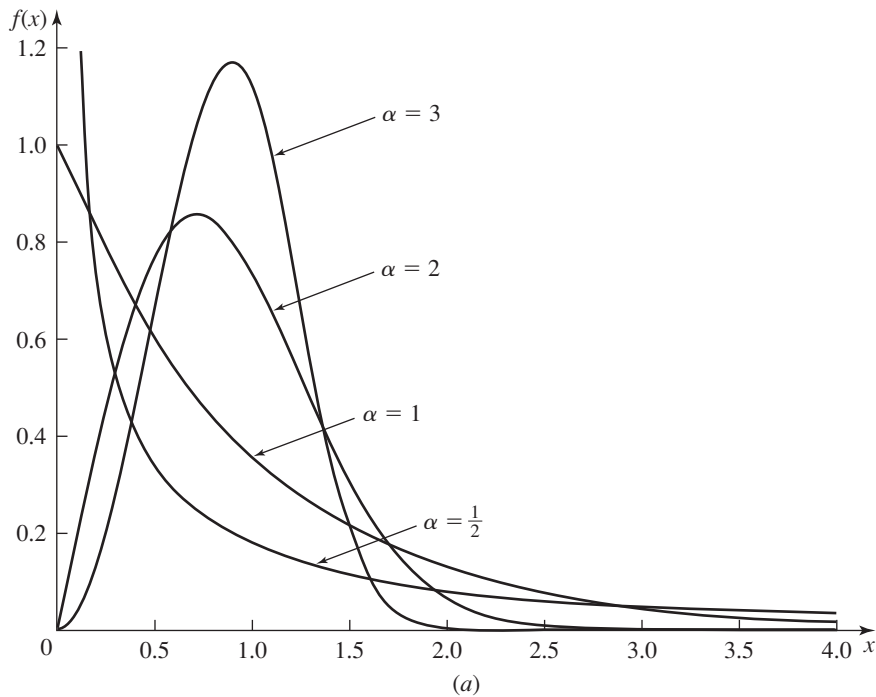


FIGURE 6.8
Weibull(α , 1) density functions. (continued)

TABLE 6.3 (continued)

Weibull	Weibull(α, β)
	where $A = \frac{\sum_{i=1}^n \ln X_i}{n}, \quad B_k = \sum_{i=1}^n X_i^{\hat{\alpha}_k}, \quad C_k = \sum_{i=1}^n X_i^{\hat{\alpha}_k} \ln X_i$ and $H_k = \sum_{i=1}^n X_i^{\hat{\alpha}_k} (\ln X_i)^2$ [See Thoman, Bain, and Antle (1969) for these formulas, as well as for confidence intervals on the true α and β .] As a starting point for the iterations, the estimate $\hat{\alpha}_0 = \left\{ \frac{(6/\pi^2) \left[\sum_{i=1}^n (\ln X_i)^2 - \left(\sum_{i=1}^n \ln X_i \right)^2 / n \right]}{n - 1} \right\}^{-1/2}$ [due to Menon (1963) and suggested in Thoman, Bain, and Antle (1969)] may be used. With this choice of $\hat{\alpha}_0$, it was reported in Thoman, Bain, and Antle (1969) that an average of only 3.5 Newton iterations were needed to achieve four-place accuracy.
Comments	1. The $\text{expo}(\beta)$ and Weibull($1, \beta$) distributions are the same. 2. $X \sim \text{Weibull}(\alpha, \beta)$ if and only if $X^\alpha \sim \text{expo}(\beta^\alpha)$ (see Prob. 6.2). 3. The (natural) logarithm of a Weibull random variable has a distribution known as the <i>extreme-value</i> or <i>Gumbel distribution</i> [see Averill M. Law & Associates (2013), Lawless (2003), and Prob. 8.1(b)]. 4. The Weibull($2, \beta$) distribution is also called a <i>Rayleigh distribution</i> with parameter β, denoted $\text{Rayleigh}(\beta)$. If Y and Z are independent normal random variables with mean 0 and variance β^2 (see the normal distribution), then $X = (Y^2 + Z^2)^{1/2} \sim \text{Rayleigh}(2^{1/2}\beta)$. 5. As $\alpha \rightarrow \infty$, the Weibull distribution becomes degenerate at β. Thus, Weibull densities for large α have a sharp peak at the mode. 6. The Weibull distribution has a negative skewness when $\alpha > 3.6$ [see Fig. 6.8(b)]. 7.$\lim_{x \rightarrow 0} f(x) = \begin{cases} \infty & \text{if } \alpha < 1 \\ \frac{1}{\beta} & \text{if } \alpha = 1 \\ 0 & \text{if } \alpha > 1 \end{cases}$
Normal	$N(\mu, \sigma^2)$
Possible applications	Errors of various types, e.g., in the impact point of a bomb; quantities that are the sum of a large number of other quantities (by virtue of central limit theorems)
Density (see Fig. 6.9)	$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)} \quad \text{for all real numbers } x$
Distribution	No closed form
Parameters	Location parameter $\mu \in (-\infty, \infty)$, scale parameter $\sigma > 0$
Range	$(-\infty, \infty)$

TABLE 6.3 (continued)

Normal	$N(\mu, \sigma^2)$
Mean	μ
Variance	σ^2
Mode	μ
MLE	$\hat{\mu} = \bar{X}(n), \quad \hat{\sigma} = \left[\frac{n-1}{n} S^2(n) \right]^{1/2}$
Comments	1. If two jointly distributed normal random variables are uncorrelated, they are also independent. For distributions other than normal, this implication is not true in general. 2. Suppose that the joint distribution of $X_1, X_2, \dots, X_m$ is multivariate normal and let $\mu_i = E(X_i)$ and $C_{ij} = \text{Cov}(X_i, X_j)$. Then for any real numbers $a, b_1, b_2, \dots, b_m$, the random variable $a + b_1X_1 + b_2X_2 + \dots + b_mX_m$ has a normal distribution with mean $\mu = a + \sum_{i=1}^m b_i\mu_i$ and variance$\sigma^2 = \sum_{i=1}^m \sum_{j=1}^m b_i b_j C_{ij}$ Note that we need <i>not</i> assume independence of the X_i's. If the X_i's are independent, then$\sigma^2 = \sum_{i=1}^m b_i^2 \text{Var}(X_i)$ 3. The $N(0, 1)$ distribution is often called the <i>standard</i> or <i>unit normal distribution</i>. 4. If $X_1, X_2, \dots, X_k$ are independent standard normal random variables, then $X_1^2 + X_2^2 + \dots + X_k^2$ has a chi-square distribution with k df, which is also the gamma($k/2, 2$) distribution.

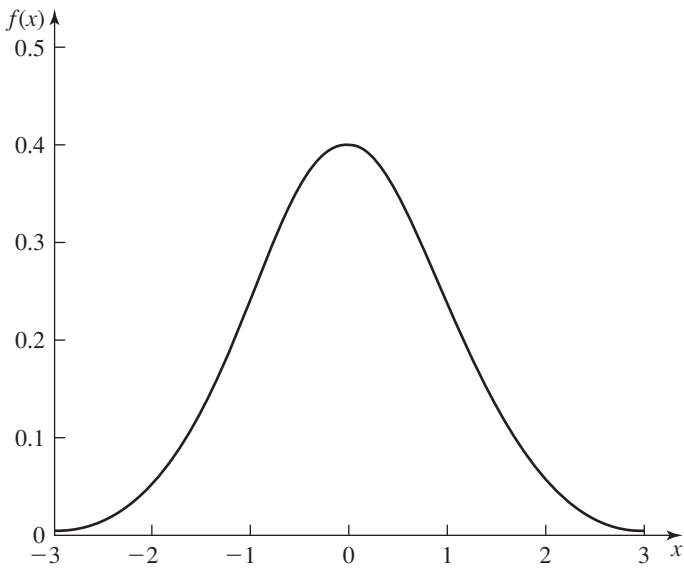


FIGURE 6.9
N(0, 1) density function. (continued)

TABLE 6.3 (continued)

Normal	$N(\mu, \sigma^2)$
	5. If $X \sim N(\mu, \sigma^2)$, then e^X has the <i>lognormal distribution</i> with scale parameter e^μ and shape parameter σ, denoted $LN(\mu, \sigma^2)$. 6. If $X \sim N(0, 1)$, if Y has a chi-square distribution with k df, and if X and Y are independent, then $X/\sqrt{Y/k}$ has a t distribution with k df (sometimes called <i>Student's t distribution</i>). 7. If the normal distribution is used to represent a nonnegative quantity (e.g., time), then its density should be truncated at $x = 0$ (see Sec. 6.8). 8. As $\sigma \rightarrow 0$, the normal distribution becomes degenerate at μ.

Lognormal	$LN(\mu, \sigma^2)$
Possible applications	Time to perform some task [density takes on shapes similar to $\text{gamma}(\alpha, \beta)$ and $\text{Weibull}(\alpha, \beta)$ densities for $\alpha > 1$, but can have a large “spike” close to $x = 0$ that is often useful]; quantities that are the product of a large number of other quantities (by virtue of central limit theorem); used as a rough model in the absence of data (see Sec. 6.11)
Density (see Fig. 6.10)	$f(x) = \begin{cases} \frac{1}{x\sqrt{2\pi\sigma^2}} \exp \frac{-(\ln x - \mu)^2}{2\sigma^2} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$
Distribution	No closed form
Parameters	Shape parameter $\sigma > 0$, scale parameter $e^\mu > 0$
Range	$[0, \infty)$
Mean	$e^{\mu+\sigma^2/2}$
Variance	$e^{2\mu+2\sigma^2}(e^{\sigma^2} - 1)$

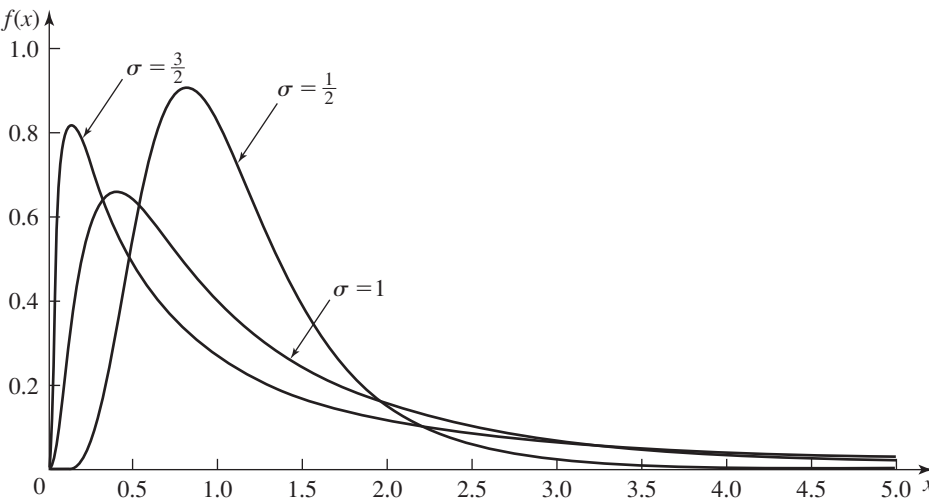


FIGURE 6.10
 $LN(0, \sigma^2)$ density functions.

TABLE 6.3 (continued)

Lognormal	$\text{LN}(\mu, \sigma^2)$
Mode	$e^{\mu - \sigma^2}$
MLE	$\hat{\mu} = \frac{\sum_{i=1}^n \ln X_i}{n}, \hat{\sigma} = \left[\frac{\sum_{i=1}^n (\ln X_i - \hat{\mu})^2}{n} \right]^{1/2}$, MLE for scale parameter = $e^{\hat{\mu}}$
Comments	1. $X \sim \text{LN}(\mu, \sigma^2)$ if and only if $\ln X \sim \text{N}(\mu, \sigma^2)$. Thus, if one has data $X_1, X_2, \dots, X_n$ that are thought to be lognormal, the logarithms of the data points $\ln X_1, \ln X_2, \dots, \ln X_n$ can be treated as normally distributed data for purposes of hypothesizing a distribution, parameter estimation, and goodness-of-fit testing. 2. As $\sigma \rightarrow 0$, the lognormal distribution becomes degenerate at e^{μ}. Thus, lognormal densities for small σ have a sharp peak at the mode. 3. $\lim_{x \rightarrow 0} f(x) = 0$, regardless of the parameter values.
Beta	$\text{beta}(\alpha_1, \alpha_2)$
Possible applications	Used as a rough model in the absence of data (see Sec. 6.11); distribution of a random proportion, such as the proportion of defective items in a shipment; time to complete a task, e.g., in a PERT network
Density (see Fig. 6.11)	$f(x) = \begin{cases} \frac{x^{\alpha_1-1}(1-x)^{\alpha_2-1}}{B(\alpha_1, \alpha_2)} & \text{if } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$ where $B(\alpha_1, \alpha_2)$ is the <i>beta function</i>, defined by $B(z_1, z_2) = \int_0^1 t^{z_1-1}(1-t)^{z_2-1} dt$ for any real numbers $z_1 > 0$ and $z_2 > 0$. Some properties of the beta function: $B(z_1, z_2) = B(z_2, z_1), \quad B(z_1, z_2) = \frac{\Gamma(z_1)\Gamma(z_2)}{\Gamma(z_1 + z_2)}$
Distribution	No closed form, in general. If either α_1 or α_2 is a positive integer, a binomial expansion can be used to obtain $F(x)$, which will be a polynomial in x , and the powers of x will be, in general, positive real numbers ranging from 0 through $\alpha_1 + \alpha_2 - 1$.
Parameters	Shape parameters $\alpha_1 > 0$ and $\alpha_2 > 0$
Range	$[0, 1]$
Mean	$\frac{\alpha_1}{\alpha_1 + \alpha_2}$
Variance	$\frac{\alpha_1 \alpha_2}{(\alpha_1 + \alpha_2)^2(\alpha_1 + \alpha_2 + 1)}$
Mode	$\begin{cases} \frac{\alpha_1 - 1}{\alpha_1 + \alpha_2 - 2} & \text{if } \alpha_1 > 1, \alpha_2 > 1 \\ 0 \text{ and } 1 & \text{if } \alpha_1 < 1, \alpha_2 < 1 \\ 0 & \text{if } (\alpha_1 < 1, \alpha_2 \geq 1) \text{ or } \text{if } (\alpha_1 = 1, \alpha_2 > 1) \\ 1 & \text{if } (\alpha_1 \geq 1, \alpha_2 < 1) \text{ or } \text{if } (\alpha_1 > 1, \alpha_2 = 1) \\ \text{does not uniquely exist} & \text{if } \alpha_1 = \alpha_2 = 1 \end{cases}$

(continued)

TABLE 6.3 (continued)

Beta	$\text{beta}(\alpha_1, \alpha_2)$
MLE	The following two equations must be satisfied: $\Psi(\hat{\alpha}_1) - \Psi(\hat{\alpha}_1 + \hat{\alpha}_2) = \ln G_1, \quad \Psi(\hat{\alpha}_2) - \Psi(\hat{\alpha}_1 + \hat{\alpha}_2) = \ln G_2$ where Ψ is the digamma function, $G_1 = (\prod_{i=1}^n X_i)^{1/n}$, and $G_2 = [\prod_{i=1}^n (1 - X_i)]^{1/n}$ [see Gnanadesikan, Pinkham, and Hughes (1967)]; note that $G_1 + G_2 \leq 1$. These equations could be solved numerically [see Beckman and Tietjen (1978)], or approximations to $\hat{\alpha}_1$ and $\hat{\alpha}_2$ can be obtained from Table 6.22 (see App. 6A), which was computed for particular (G_1, G_2) pairs by modifications of the methods in Beckman and Tietjen (1978).
Comments	1. The $U(0, 1)$ and $\text{beta}(1, 1)$ distributions are the same. 2. If X_1 and X_2 are independent random variables with $X_i \sim \text{gamma}(\alpha_i, \beta)$, then $X_1/(X_1 + X_2) \sim \text{beta}(\alpha_1, \alpha_2)$. 3. A beta random variable X on $[0, 1]$ can be rescaled and relocated to obtain a beta random variable on $[a, b]$ of the same shape by the transformation $a + (b - a)X$. 4. $X \sim \text{beta}(\alpha_1, \alpha_2)$ if and only if $1 - X \sim \text{beta}(\alpha_2, \alpha_1)$. 5. $X \sim \text{beta}(\alpha_1, \alpha_2)$ if and only if $Y = X/(1 - X)$ has a Pearson type VI distribution with shape parameters α_1, α_2 and scale parameter 1, denoted $\text{PT6}(\alpha_1, \alpha_2, 1)$.

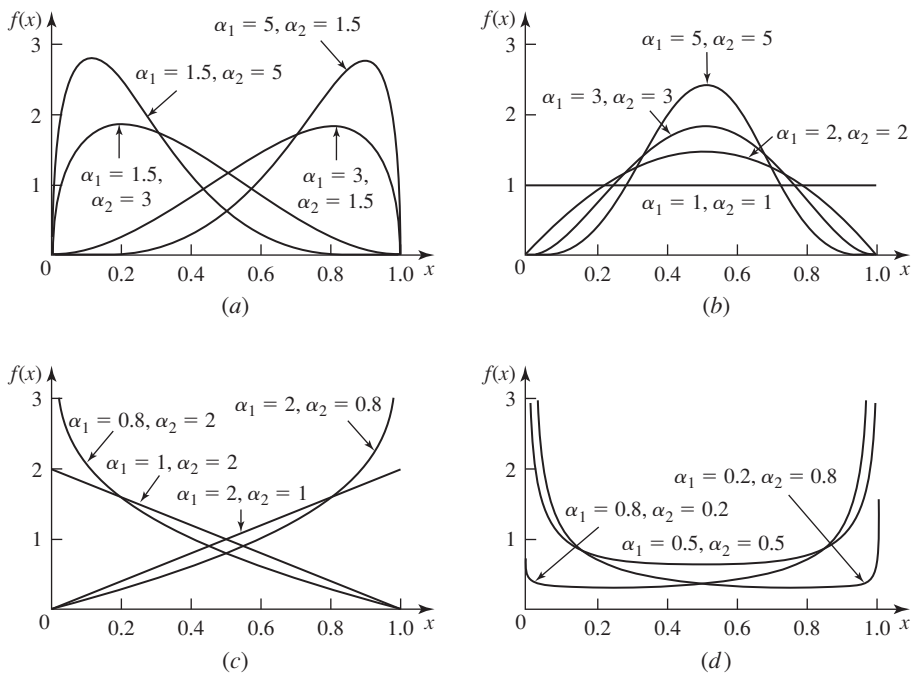


FIGURE 6.11
 $\text{beta}(\alpha_1, \alpha_2)$ density functions.

TABLE 6.3 (continued)

Beta	beta(α_1, α_2)
	6. The beta(1, 2) density is a left triangle, and the beta(2, 1) density is a right triangle.
	7.
	$\lim_{x \rightarrow 0} f(x) = \begin{cases} \infty & \text{if } \alpha_1 < 1 \\ \alpha_2 & \text{if } \alpha_1 = 1, \\ 0 & \text{if } \alpha_1 > 1 \end{cases} \quad \lim_{x \rightarrow 1} f(x) = \begin{cases} \infty & \text{if } \alpha_2 < 1 \\ \alpha_1 & \text{if } \alpha_2 = 1 \\ 0 & \text{if } \alpha_2 > 1 \end{cases}$
	8. The density is symmetric about $x = \frac{1}{2}$ if and only if $\alpha_1 = \alpha_2$. Also, the mean and the mode are equal if and only if $\alpha_1 = \alpha_2 > 1$.

Pearson type V	PT5(α, β)
Possible applications	Time to perform some task (density takes on shapes similar to lognormal, but can have a larger “spike” close to $x = 0$)
Density (see Fig. 6.12)	$f(x) = \begin{cases} \frac{x^{-(\alpha+1)}e^{-\beta/x}}{\beta^{-\alpha}\Gamma(\alpha)} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$
Distribution	$F(x) = \begin{cases} 1 - F_G\left(\frac{1}{x}\right) & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$ where $F_G(x)$ is the distribution function of a gamma($\alpha, 1/\beta$) random variable
Parameters	Shape parameter $\alpha > 0$, scale parameter $\beta > 0$

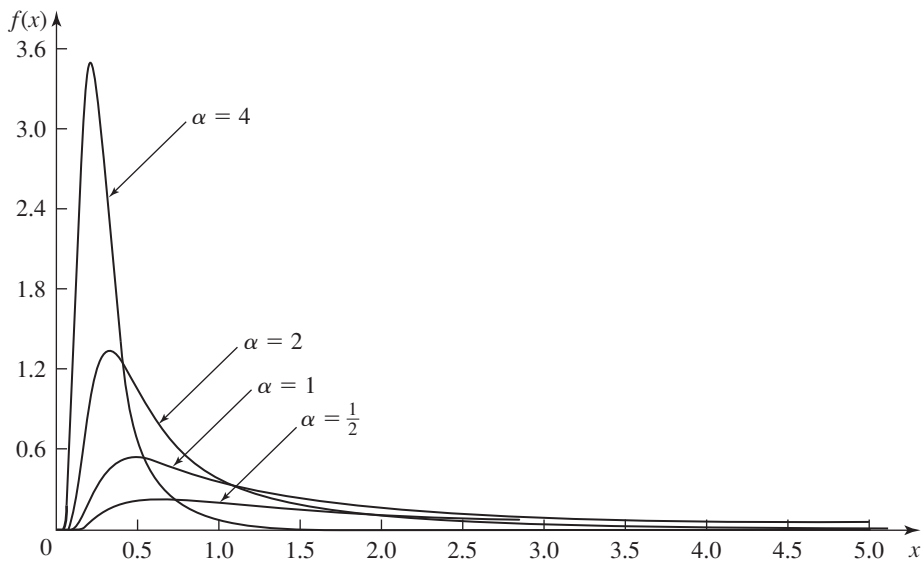


FIGURE 6.12
PT5($\alpha, 1$) density functions.

(continued)

TABLE 6.3 (continued)

Pearson type V	PT5(α, β)
Range	$[0, \infty)$
Mean	$\frac{\beta}{\alpha - 1}$ for $\alpha > 1$
Variance	$\frac{\beta^2}{(\alpha - 1)^2(\alpha - 2)}$ for $\alpha > 2$
Mode	$\frac{\beta}{\alpha + 1}$
MLE	If one has data $X_1, X_2, \dots, X_n$, then fit a $\text{gamma}(\alpha_G, \beta_G)$ distribution to $1/X_1, 1/X_2, \dots, 1/X_n$, resulting in the maximum-likelihood estimators $\hat{\alpha}_G$ and $\hat{\beta}_G$. Then the maximum-likelihood estimators for the PT5(α, β) are $\hat{\alpha} = \hat{\alpha}_G$ and $\hat{\beta} = 1/\hat{\beta}_G$ (see comment 1 below).
Comments	1. $X \sim \text{PT5}(\alpha, \beta)$ if and only if $Y = 1/X \sim \text{gamma}(\alpha, 1/\beta)$. Thus, the Pearson type V distribution is sometimes called the <i>inverted gamma distribution</i>. 2. Note that the mean and variance exist only for certain values of the shape parameter.

Pearson type VI	PT6($\alpha_1, \alpha_2, \beta$)
Possible applications	Time to perform some task
Density (see Fig. 6.13)	$f(x) = \begin{cases} \frac{(x/\beta)^{\alpha_1 - 1}}{\beta B(\alpha_1, \alpha_2)[1 + (x/\beta)]^{\alpha_1 + \alpha_2}} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$
Distribution	$F(x) = \begin{cases} F_B\left(\frac{x}{x + \beta}\right) & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$ where $F_B(x)$ is the distribution function of a $\text{beta}(\alpha_1, \alpha_2)$ random variable
Parameters	Shape parameters $\alpha_1 > 0$ and $\alpha_2 > 0$, scale parameter $\beta > 0$
Range	$[0, \infty)$
Mean	$\frac{\beta\alpha_1}{\alpha_2 - 1}$ for $\alpha_2 > 1$
Variance	$\frac{\beta^2\alpha_1(\alpha_1 + \alpha_2 - 1)}{(\alpha_2 - 1)^2(\alpha_2 - 2)}$ for $\alpha_2 > 2$
Mode	$\begin{cases} \frac{\beta(\alpha_1 - 1)}{\alpha_2 + 1} & \text{if } \alpha_1 \geq 1 \\ 0 & \text{otherwise} \end{cases}$
MLE	If one has data $X_1, X_2, \dots, X_n$ that are thought to be PT6($\alpha_1, \alpha_2, 1$), then fit a $\text{beta}(\alpha_1, \alpha_2)$ distribution to $X_i/(1 + X_i)$ for $i = 1, 2, \dots, n$, resulting in the maximum-likelihood estimators $\hat{\alpha}_1$ and $\hat{\alpha}_2$. Then the maximum-likelihood estimators for the PT6($\alpha_1, \alpha_2, 1$) (note that $\beta = 1$) distribution are also $\hat{\alpha}_1$ and $\hat{\alpha}_2$ (see comment 1 below).

TABLE 6.3 (continued)

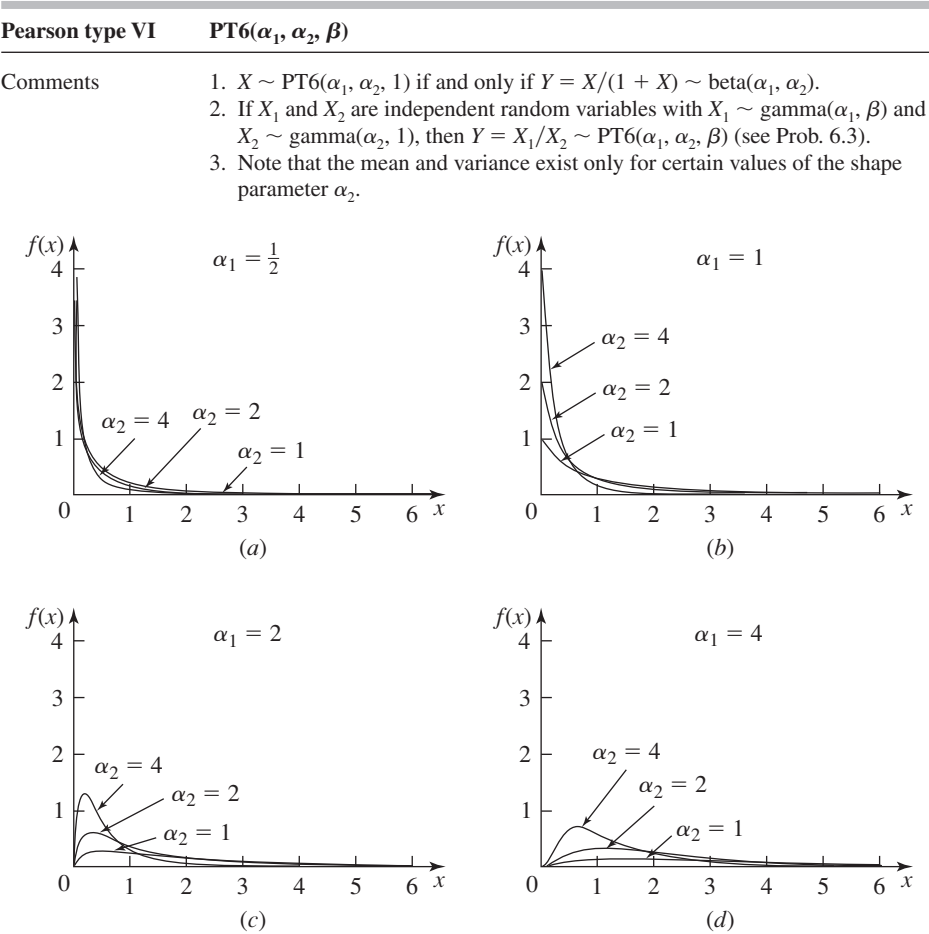


FIGURE 6.13
PT6($\alpha_1, \alpha_2, 1$) density functions.

Log-logistic	LL(α, β)
Possible applications	Time to perform some task
Density (see Fig. 6.14)	$f(x) = \begin{cases} \frac{\alpha(x/\beta)^{\alpha-1}}{\beta[1 + (x/\beta)^\alpha]^2} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$
Distribution	$F(x) = \begin{cases} \frac{1}{1 + (x/\beta)^{-\alpha}} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$
Parameters	Shape parameter $\alpha > 0$, scale parameter $\beta > 0$

(continued)

TABLE 6.3 (continued)

Log-logistic	LL(α, β)
Range	$[0, \infty)$
Mean	$\beta\theta \operatorname{cosecant}(\theta)$ for $\alpha > 1$, where $\theta = \pi/\alpha$
Variance	$\beta^2\theta\{2 \operatorname{cosecant}(2\theta) - \theta[\operatorname{cosecant}(\theta)]^2\}$ for $\alpha > 2$
Mode	$\begin{cases} \beta\left(\frac{\alpha-1}{\alpha+1}\right)^{1/\alpha} & \text{if } \alpha > 1 \\ 0 & \text{otherwise} \end{cases}$
MLE	Let $Y_i = \ln X_i$. Solve the following two equations for $\hat{a}$ and $\hat{b}$: $\sum_{i=1}^n [1 + e^{(Y_i - \hat{a})/\hat{b}}]^{-1} = \frac{n}{2} \tag{6.1}$

and

$$\sum_{i=1}^n \left(\frac{Y_i - \hat{a}}{\hat{b}}\right) \frac{1 - e^{(Y_i - \hat{a})/\hat{b}}}{1 + e^{(Y_i - \hat{a})/\hat{b}}} = n \tag{6.2}$$

Then the MLEs for the log-logistic distribution are $\hat{\alpha} = 1/\hat{a}$ and $\hat{\beta} = e^{\hat{b}}$. Johnson, Kotz, and Balakrishnan (1995, chap. 23) suggest solving Eqs. (6.1) and (6.2) by using Newton's method.

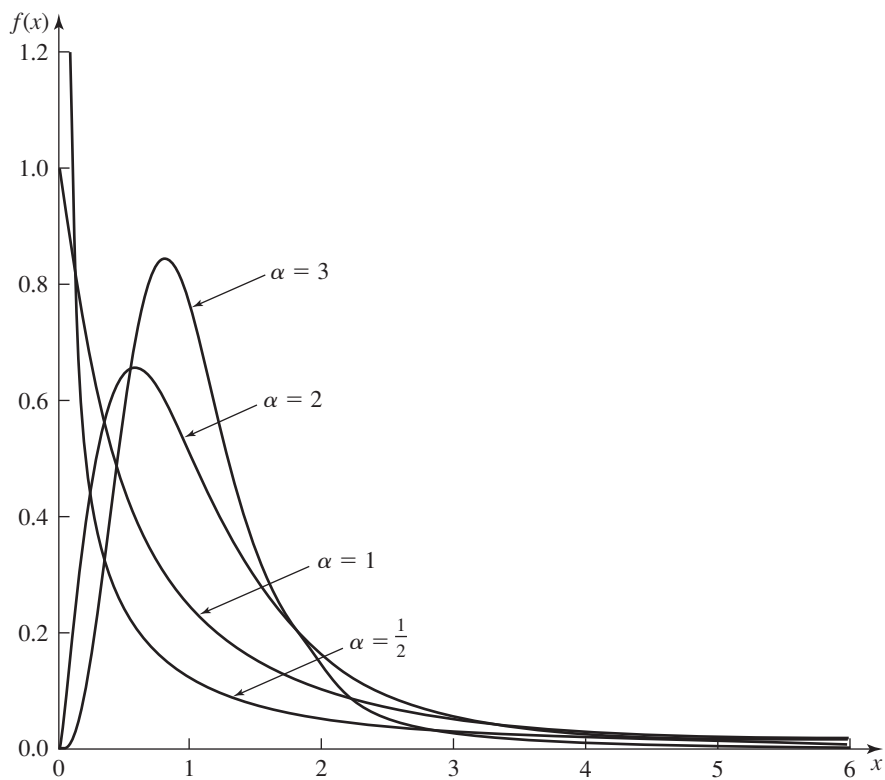


FIGURE 6.14
LL($\alpha, 1$) density functions.

TABLE 6.3 (continued)

Log-logistic	$LL(\alpha, \beta)$
Comment	$X \sim LL(\alpha, \beta)$ if and only if $\ln X$ is distributed as a logistic distribution (see Prob. 8.1) with location parameter $\ln \beta$ and scale parameter $1/\alpha$. Thus, if one has data $X_1, X_2, \dots, X_n$ that are thought to be log-logistic, the logarithms of the data points $\ln X_1, \ln X_2, \dots, \ln X_n$ can be treated as having a logistic distribution for purposes of hypothesizing a distribution, parameter estimation, and goodness-of-fit testing.
Johnson S_B	$JSB(\alpha_1, \alpha_2, a, b)$
Density (see Fig. 6.15)	$f(x) = \begin{cases} \frac{\alpha_2(b-a)}{(x-a)(b-x)\sqrt{2\pi}} e^{-\frac{1}{2}\left[\alpha_1 + \alpha_2 \ln\left(\frac{x-a}{b-x}\right)\right]^2} & \text{if } a < x < b \\ 0 & \text{otherwise} \end{cases}$
Distribution	$F(x) = \begin{cases} \Phi\left[\alpha_1 + \alpha_2 \ln\left(\frac{x-a}{b-x}\right)\right] & \text{if } a < x < b \\ 0 & \text{otherwise} \end{cases}$ where $\Phi(x)$ is the distribution function of a normal random variable with $\mu = 0$ and $\sigma^2 = 1$
Parameters	Location parameter $a \in (-\infty, \infty)$, scale parameter $b - a$ ($b > a$), shape parameters $\alpha_1 \in (-\infty, \infty)$ and $\alpha_2 > 0$
Range	$[a, b]$
Mean	All moments exist but are extremely complicated [see Johnson, Kotz, and Balakrishnan (1994, p. 35)]
Mode	The density is bimodal when $\alpha_2 < \frac{1}{\sqrt{2}}$ and $ \alpha_1 < \frac{\sqrt{1 - 2\alpha_2^2}}{\alpha_2} - 2\alpha_2 \tanh^{-1}(\sqrt{1 - 2\alpha_2^2})$ [$\tanh^{-1}$ is the inverse hyperbolic tangent]; otherwise the density is unimodal. The equation satisfied by any mode x, other than at the endpoints of the range, is $\frac{2(x-a)}{b-a} = 1 + \alpha_1\alpha_2 + \alpha_2^2 \ln\left(\frac{x-a}{b-x}\right)$
Comments	1. $X \sim JSB(\alpha_1, \alpha_2, a, b)$ if and only if $\alpha_1 + \alpha_2 \ln\left(\frac{X-a}{b-X}\right) \sim N(0, 1)$ 2. The density function is skewed to the left, symmetric, or skewed to the right if $\alpha_1 > 0$, $\alpha_1 = 0$, or $\alpha_1 < 0$, respectively. 3. $\lim_{x \rightarrow a} f(x) = \lim_{x \rightarrow b} f(x) = 0$ for all values of α_1 and α_2. 4. The four parameters may be estimated using a number of methods [see, for example, Swain, Venkatraman, and Wilson (1988) and Slifker and Shapiro (1980)].

(continued)

TABLE 6.3 (continued)

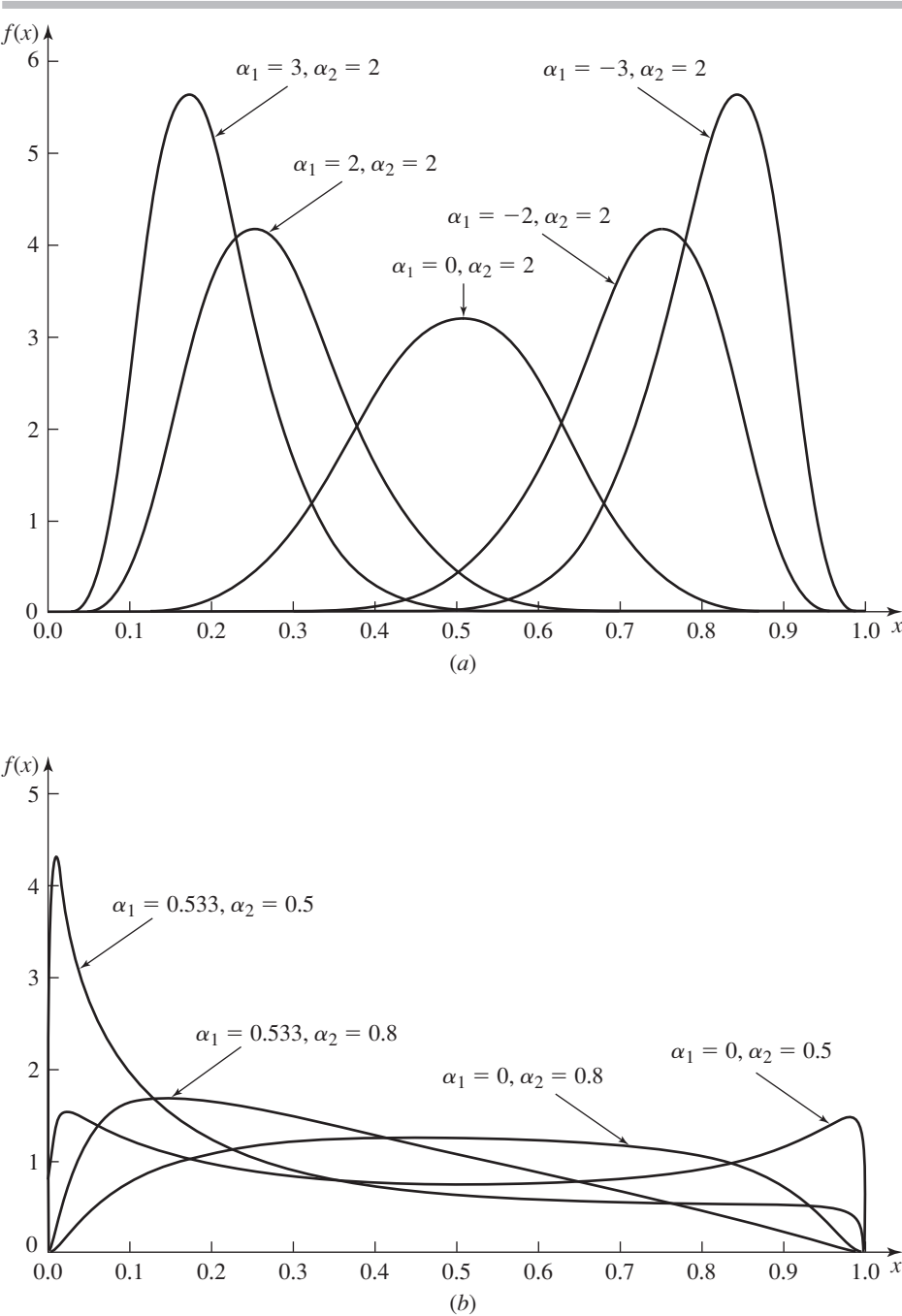


FIGURE 6.15
JSB($\alpha_1, \alpha_2, 0, 1$) density functions.

TABLE 6.3 (continued)

Johnson S_U	$\text{JSU}(\alpha_1, \alpha_2, \gamma, \beta)$
Density (see Fig. 6.16)	$f(x) = \frac{\alpha_2}{\sqrt{2\pi}\sqrt{(x-\gamma)^2 + \beta^2}} e^{-\frac{1}{2}\left\{\alpha_1 + \alpha_2 \ln\left[\frac{x-\gamma}{\beta} + \sqrt{\left(\frac{x-\gamma}{\beta}\right)^2 + 1}\right]\right\}^2} \quad \text{for } -\infty < x < \infty$
Distribution	$F(x) = \Phi\left\{\alpha_1 + \alpha_2 \ln\left[\frac{x-\gamma}{\beta} + \sqrt{\left(\frac{x-\gamma}{\beta}\right)^2 + 1}\right]\right\} \quad \text{for } -\infty < x < \infty$
Parameters	Location parameter $\gamma \in (-\infty, \infty)$, scale parameter $\beta > 0$, shape parameters $\alpha_1 \in (-\infty, \infty)$ and $\alpha_2 > 0$
Range	$(-\infty, \infty)$
Mean	$\gamma - \beta e^{1/(2\alpha_2^2)} \sinh\left(\frac{\alpha_1}{\alpha_2}\right)$, where $\sinh$ is the hyperbolic sine
Mode	The equation satisfied by the mode, other than at the endpoints of the range, is $\gamma + \beta y$, where y satisfies $y + \alpha_1\alpha_2\sqrt{y^2 + 1} + \alpha_2^2\sqrt{y^2 + 1} \ln(y + \sqrt{y^2 + 1}) = 0$
Comments	1. $X \sim \text{JSU}(\alpha_1, \alpha_2, \gamma, \beta)$ if and only if$\alpha_1 + \alpha_2 \ln\left[\frac{X-\gamma}{\beta} + \sqrt{\left(\frac{X-\gamma}{\beta}\right)^2 + 1}\right] \sim N(0, 1)$ 2. The density function is skewed to the left, symmetric, or skewed to the right if $\alpha_1 > 0$, $\alpha_1 = 0$, or $\alpha_1 < 0$, respectively. 3. The four parameters may be estimated by a number of methods [see, for example, Swain, Venkatraman, and Wilson (1988) and Slifker and Shapiro (1980)].

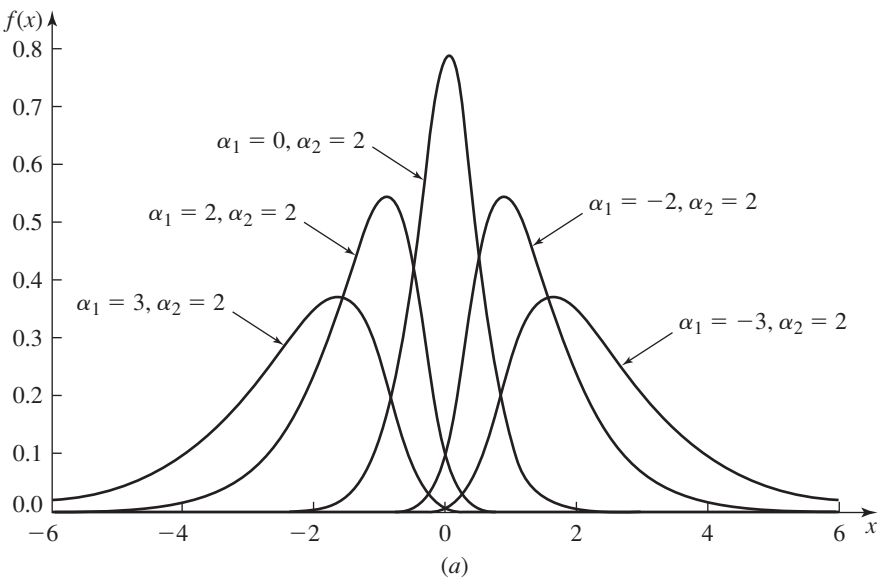


FIGURE 6.16
 $\text{JSU}(\alpha_1, \alpha_2, 0, 1)$ density functions.

(continued)

TABLE 6.3 (continued)

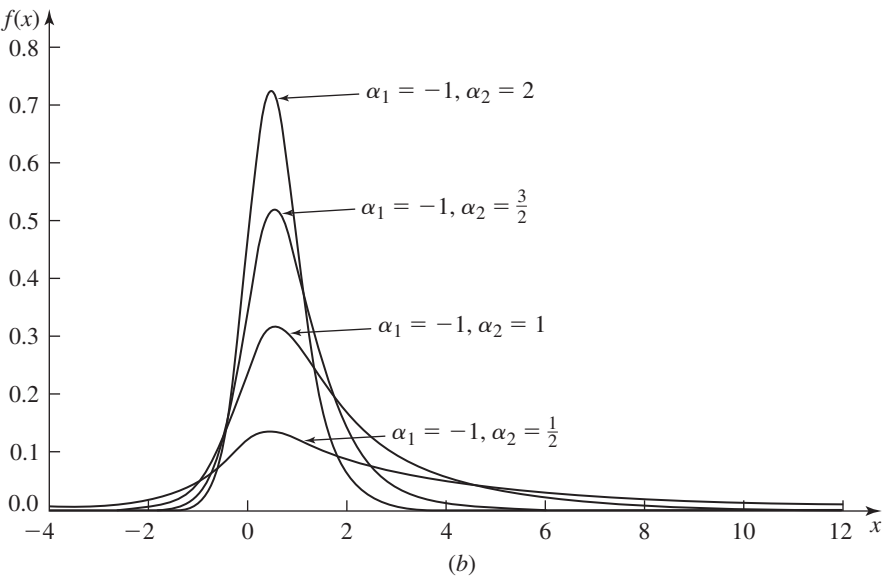


FIGURE 6.16
(continued)

Triangular	$\text{triang}(a, b, m)$
Possible applications	Used as a rough model in the absence of data (see Sec. 6.11)
Density (see Fig. 6.17)	$f(x) = \begin{cases} \frac{2(x-a)}{(b-a)(m-a)} & \text{if } a \leq x \leq m \\ \frac{2(b-x)}{(b-a)(b-m)} & \text{if } m < x \leq b \\ 0 & \text{otherwise} \end{cases}$
Distribution	$F(x) = \begin{cases} 0 & \text{if } x < a \\ \frac{(x-a)^2}{(b-a)(m-a)} & \text{if } a \leq x \leq m \\ 1 - \frac{(b-x)^2}{(b-a)(b-m)} & \text{if } m < x \leq b \\ 1 & \text{if } b < x \end{cases}$
Parameters	a, b , and m real numbers with $a < m < b$. a is a location parameter, $b - a$ is a scale parameter, m is a shape parameter
Range	$[a, b]$
Mean	$\frac{a + b + m}{3}$

TABLE 6.3 (continued)

Triangular	$\text{triang}(a, b, m)$
Variance	$\frac{a^2 + b^2 + m^2 - ab - am - bm}{18}$
Mode	m
MLE	Our use of the triangular distribution, as described in Sec. 6.11, is as a rough model when there are no data. Thus, MLEs are not relevant.
Comment	The limiting cases as $m \rightarrow b$ and $m \rightarrow a$ are called the <i>right triangular</i> and <i>left triangular distributions</i> , respectively, and are discussed in Prob. 8.7. For $a = 0$ and $b = 1$, both the left and right triangular distributions are special cases of the beta distribution.

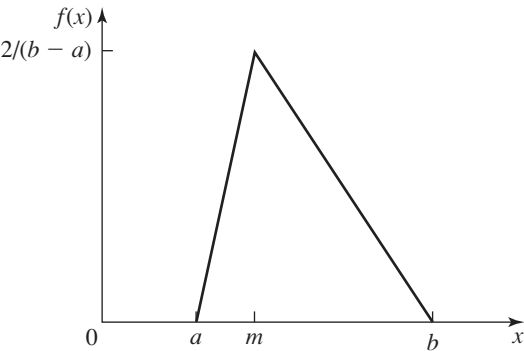


FIGURE 6.17
 $\text{triang}(a, b, m)$ density function.

6.2.3 Discrete Distributions

The descriptions of the six discrete distributions in Table 6.4 follow the same pattern as for the continuous distributions in Table 6.3.

6.2.4 Empirical Distributions

In some situations we might want to use the observed data themselves to specify directly (in some sense) a distribution, called an *empirical distribution*, from which random values are generated during the simulation, rather than fitting a theoretical distribution to the data. For example, it could happen that we simply cannot find a theoretical distribution that fits the data adequately (see Secs. 6.4 through 6.6). This section explores ways of specifying empirical distributions.

For continuous random variables, the type of empirical distribution that can be defined depends on whether we have the actual values of the individual original observations $X_1, X_2, \dots, X_n$ rather than only the *number* of X_i 's that fall into each of several specified intervals. (The latter case is called *grouped data* or *data in the form of a histogram*.) If the original data are available, we can define a continuous, piecewise-linear distribution function F by first sorting the X_i 's into increasing

TABLE 6.4
Discrete distributions

Bernoulli	Bernoulli(<i>p</i>)
Possible applications	Random occurrence with two possible outcomes; used to generate other discrete random variates, e.g., binomial, geometric, and negative binomial
Mass (see Fig. 6.18)	$p(x) = \begin{cases} 1 - p & \text{if } x = 0 \\ p & \text{if } x = 1 \\ 0 & \text{otherwise} \end{cases}$
Distribution	$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - p & \text{if } 0 \leq x < 1 \\ 1 & \text{if } 1 \leq x \end{cases}$
Parameter	$p \in (0, 1)$
Range	{0, 1}
Mean	p
Variance	$p(1 - p)$
Mode	$\begin{cases} 0 & \text{if } p < \frac{1}{2} \\ 0 \text{ and } 1 & \text{if } p = \frac{1}{2} \\ 1 & \text{if } p > \frac{1}{2} \end{cases}$
MLE	$\hat{p} = \bar{X}(n)$
Comments	1. A Bernoulli(<i>p</i>) random variable <i>X</i> can be thought of as the outcome of an experiment that either “fails” or “succeeds.” If the probability of success is <i>p</i>, and we let <i>X</i> = 0 if the experiment fails and <i>X</i> = 1 if it succeeds, then $X \sim \text{Bernoulli}(p)$. Such an experiment, often called a <i>Bernoulli trial</i>, provides a convenient way of relating several other discrete distributions to the Bernoulli distribution. 2. If <i>t</i> is a positive integer and <i>X</i>₁, <i>X</i>₂, . . . , <i>X</i>_{<i>t</i>} are independent Bernoulli(<i>p</i>) random variables, then <i>X</i>₁ + <i>X</i>₂ + . . . + <i>X</i>_{<i>t</i>} has the binomial distribution with parameters <i>t</i> and <i>p</i>. Thus, a binomial random variable can be thought of as the number of successes in a fixed number of independent Bernoulli trials. 3. Suppose we begin making independent replications of a Bernoulli trial with probability <i>p</i> of success on each trial. Then the number of failures <i>before</i> observing the first success has a geometric distribution with parameter <i>p</i>. For a positive integer <i>s</i>, the number of failures before

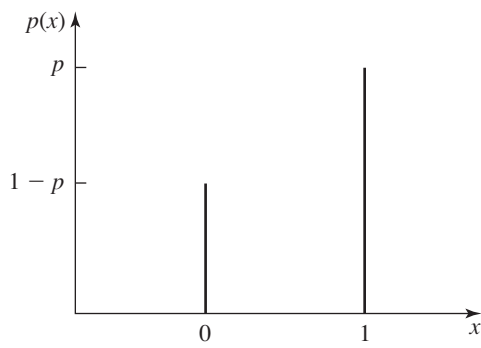


FIGURE 6.18
Bernoulli(*p*) mass function (*p* > 0.5 here).

TABLE 6.4 (continued)

Bernoulli	Bernoulli(p)
	observing the s th success has a negative binomial distribution with parameters s and p .
	4. The Bernoulli(p) distribution is a special case of the binomial distribution (with $t = 1$ and the same value for p).
Discrete uniform	DU(i, j)
Possible applications	Random occurrence with several possible outcomes, each of which is equally likely; used as a “first” model for a quantity that is varying among the integers i through j but about which little else is known
Mass (see Fig. 6.19)	$p(x) = \begin{cases} \frac{1}{j - i + 1} & \text{if } x \in \{i, i + 1, \dots, j\} \\ 0 & \text{otherwise} \end{cases}$
Distribution	$F(x) = \begin{cases} 0 & \text{if } x < i \\ \frac{\lfloor x \rfloor - i + 1}{j - i + 1} & \text{if } i \leq x \leq j \\ 1 & \text{if } j < x \end{cases}$ where $\lfloor x \rfloor$ denotes the largest integer $\leq x$
Parameters	i and j integers with $i \leq j$; i is a location parameter, $j - i$ is a scale parameter
Range	$\{i, i + 1, \dots, j\}$
Mean	$\frac{i + j}{2}$
Variance	$\frac{(j - i + 1)^2 - 1}{12}$
Mode	Does not uniquely exist
MLE	$\hat{i} = \min_{1 \leq k \leq n} X_k, \quad \hat{j} = \max_{1 \leq k \leq n} X_k$
Comment	The DU(0, 1) and Bernoulli($\frac{1}{2}$) distributions are the same.

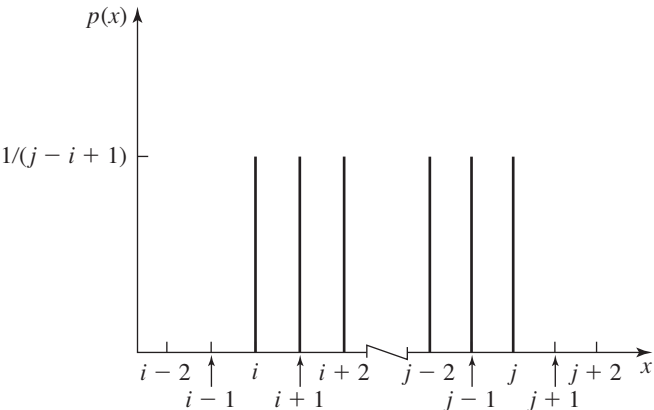


FIGURE 6.19
DU(i, j) mass function.
(continued)

TABLE 6.4 (continued)

Binomial	$\text{bin}(t, p)$
Possible applications	Number of successes in t independent Bernoulli trials with probability p of success on each trial; number of “defective” items in a batch of size t ; number of items in a batch (e.g., a group of people) of random size; number of items demanded from an inventory
Mass (see Fig. 6.20)	$p(x) = \begin{cases} \binom{t}{x} p^x (1-p)^{t-x} & \text{if } x \in \{0, 1, \dots, t\} \\ 0 & \text{otherwise} \end{cases}$ where $\binom{t}{x}$ is the <i>binomial coefficient</i>, defined by $\binom{t}{x} = \frac{t!}{x!(t-x)!}$
Distribution	$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ \sum_{i=0}^{\lfloor x \rfloor} \binom{t}{i} p^i (1-p)^{t-i} & \text{if } 0 \leq x \leq t \\ 1 & \text{if } t < x \end{cases}$
Parameters	t a positive integer, $p \in (0, 1)$
Range	$\{0, 1, \dots, t\}$

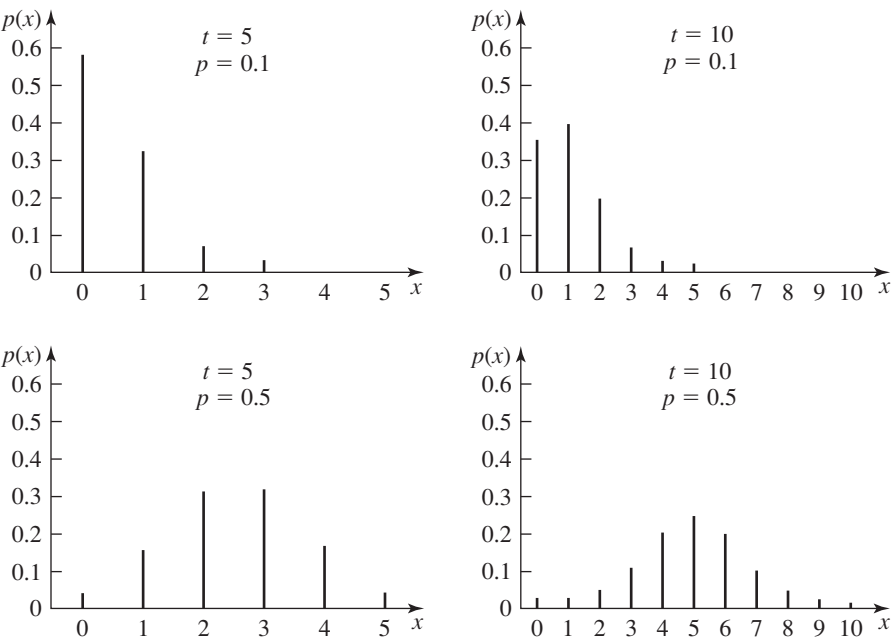


FIGURE 6.20
 $\text{bin}(t, p)$ mass functions.

TABLE 6.4 (continued)

Binomial	$\text{bin}(t, p)$
Mean	tp
Variance	$tp(1 - p)$
Mode	$\begin{cases} p(t + 1) - 1 \text{ and } p(t + 1) & \text{if } p(t + 1) \text{ is an integer} \\ \lfloor p(t + 1) \rfloor & \text{otherwise} \end{cases}$
MLE	If t is known, then $\hat{p} = \bar{X}(n)/t$. If both t and p are unknown, then $\hat{t}$ and $\hat{p}$ exist if and only if $\bar{X}(n) > (n - 1)S^2(n)/n = V(n)$. Then the following approach could be taken. Let $M = \max_{1 \leq i \leq n} X_i$, and for $k = 0, 1, \dots, M$, let f_k be the number of X_i's $\geq k$. Then it can be shown that $\hat{t}$ and $\hat{p}$ are the values for t and p that maximize the function $g(t, p) = \sum_{k=1}^M f_k \ln(t - k + 1) + nt \ln(1 - p) + n\bar{X}(n) \ln \frac{p}{1 - p}$ subject to the constraints that $t \in \{M, M + 1, \dots\}$ and $0 < p < 1$. It is easy to see that for a fixed value of t, say t_0, the value of p that maximizes $g(t_0, p)$ is $\bar{X}(n)/t_0$, so $\hat{t}$ and $\hat{p}$ are the values of t and $\bar{X}(n)/t$ that lead to the largest value of $g[t, \bar{X}(n)/t]$ for $t \in \{M, M + 1, \dots, M'\}$, where M' is given by [see DeRiggi (1983)] $M' = \left\lfloor \frac{\bar{X}(n)(M - 1)}{1 - [V(n)/\bar{X}(n)]} \right\rfloor$ Note also that $g[t, \bar{X}(n)/t]$ is a unimodal function of t.
Comments	1. If $Y_1, Y_2, \dots, Y_t$ are independent Bernoulli(p) random variables, then $Y_1 + Y_2 + \dots + Y_t \sim \text{bin}(t, p)$. 2. If $X_1, X_2, \dots, X_m$ are independent random variables and $X_i \sim \text{bin}(t_i, p)$, then $X_1 + X_2 + \dots + X_m \sim \text{bin}(t_1 + t_2 + \dots + t_m, p)$. 3. The $\text{bin}(t, p)$ distribution is symmetric if and only if $p = \frac{1}{2}$. 4. $X \sim \text{bin}(t, p)$ if and only if $t - X \sim \text{bin}(t, 1 - p)$. 5. The $\text{bin}(1, p)$ and Bernoulli(p) distributions are the same.

Geometric	$\text{geom}(p)$
Possible applications	Number of failures before the first success in a sequence of independent Bernoulli trials with probability p of success on each trial; number of items inspected before encountering the first defective item; number of items in a batch of random size; number of items demanded from an inventory
Mass (see Fig. 6.21)	$p(x) = \begin{cases} p(1 - p)^x & \text{if } x \in \{0, 1, \dots\} \\ 0 & \text{otherwise} \end{cases}$
Distribution	$F(x) = \begin{cases} 1 - (1 - p)^{\lfloor x \rfloor + 1} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$
Parameter	$p \in (0, 1)$
Range	$\{0, 1, \dots\}$
Mean	$\frac{1 - p}{p}$

(continued)

TABLE 6.4 (continued)

Geometric	$\text{geom}(p)$
Variance	$\frac{1-p}{p^2}$
Mode	0
MLE	$\hat{p} = \frac{1}{\bar{X}(n) + 1}$
Comments	1. If $Y_1, Y_2, \dots$ is a sequence of independent Bernoulli(p) random variables and $X = \min\{i: Y_i = 1\} - 1$, then $X \sim \text{geom}(p)$. 2. If $X_1, X_2, \dots, X_s$ are independent $\text{geom}(p)$ random variables, then $X_1 + X_2 + \dots + X_s$ has a negative binomial distribution with parameters s and p. 3. The geometric distribution is the discrete analog of the exponential distribution, in the sense that it is the only discrete distribution with the memoryless property (see Prob. 4.31). 4. The $\text{geom}(p)$ distribution is a special case of the negative binomial distribution (with $s = 1$ and the same value for p).

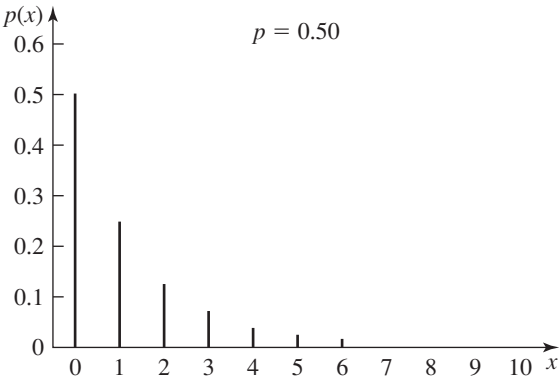
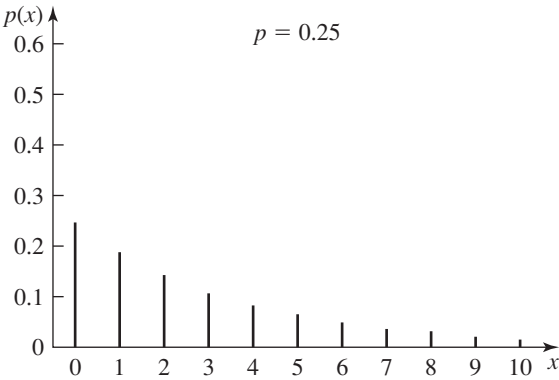


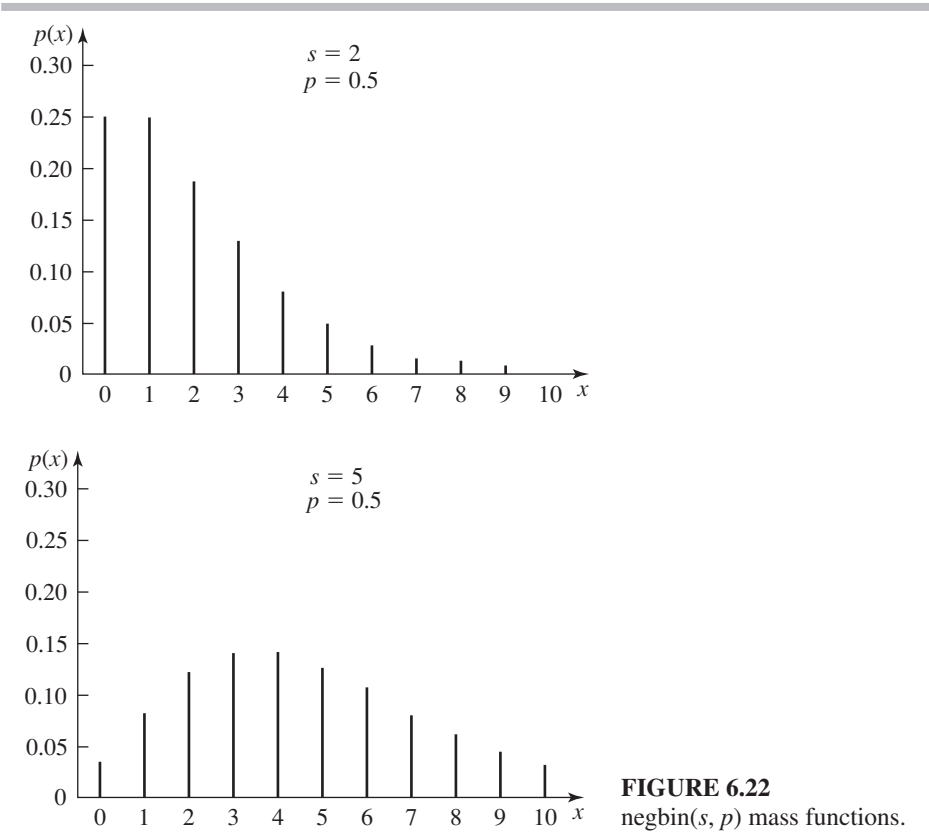
FIGURE 6.21
 $\text{geom}(p)$ mass functions.

TABLE 6.4 (continued)

Negative binomial	negbin(s, p)
Possible applications	Number of failures before the s th success in a sequence of independent Bernoulli trials with probability p of success on each trial; number of good items inspected before encountering the s th defective item; number of items in a batch of random size; number of items demanded from an inventory
Mass (see Fig. 6.22)	$p(x) = \begin{cases} \binom{s+x-1}{x} p^s (1-p)^x & \text{if } x \in \{0, 1, \dots\} \\ 0 & \text{otherwise} \end{cases}$
Distribution	$F(x) = \begin{cases} \sum_{i=0}^{\lfloor x \rfloor} \binom{s+i-1}{i} p^s (1-p)^i & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$
Parameters	s a positive integer, $p \in (0, 1)$
Range	$\{0, 1, \dots\}$
Mean	$\frac{s(1-p)}{p}$
Variance	$\frac{s(1-p)}{p^2}$
Mode	Let $y = [s(1-p) - 1]/p$; then $\text{Mode} = \begin{cases} y \text{ and } y + 1 & \text{if } y \text{ is an integer} \\ \lfloor y \rfloor + 1 & \text{otherwise} \end{cases}$
MLE	If s is known, then $\hat{p} = s/[\bar{X}(n) + s]$. If both s and p are unknown, then $\hat{s}$ and $\hat{p}$ exist if and only if $V(n) = (n-1)S^2(n)/n > \bar{X}(n)$. Let $M = \max_{1 \leq i \leq n} X_i$, and for $k = 0, 1, \dots, M$, let f_k be the number of X_i's $\geq k$. Then we can show that $\hat{s}$ and $\hat{p}$ are the values for s and p that maximize the function $h(s, p) = \sum_{k=1}^M f_k \ln(s+k-1) + ns \ln p + n\bar{X}(n) \ln(1-p)$ subject to the constraints that $s \in \{1, 2, \dots\}$ and $0 < p < 1$. For a fixed value of s, say s_0, the value of p that maximizes $h(s_0, p)$ is $s_0/[\bar{X}(n) + s_0]$, so that we could examine $h(1, 1/[\bar{X}(n) + 1])$, $h(2, 2/[\bar{X}(n) + 2])$, $\dots$. Then $\hat{s}$ and $\hat{p}$ are chosen to be the values of s and $s/[\bar{X}(n) + s]$ that lead to the biggest observed value of $h(s, s/[\bar{X}(n) + s])$. However, since $h(s, s/[\bar{X}(n) + s])$ is a unimodal function of s [see Levin and Reeds (1977)], it is clear when to terminate the search.
Comments	1. If $Y_1, Y_2, \dots, Y_s$ are independent $\text{geom}(p)$ random variables, then $Y_1 + Y_2 + \dots + Y_s \sim \text{negbin}(s, p)$. 2. If $Y_1, Y_2, \dots$ is a sequence of independent $\text{Bernoulli}(p)$ random variables and $X = \min\{i: \sum_{j=1}^i Y_j = s\} - s$, then $X \sim \text{negbin}(s, p)$. 3. If $X_1, X_2, \dots, X_m$ are independent random variables and $X_i \sim \text{negbin}(s_i, p)$, then $X_1 + X_2 + \dots + X_m \sim \text{negbin}(s_1 + s_2 + \dots + s_m, p)$. 4. The $\text{negbin}(1, p)$ and $\text{geom}(p)$ distributions are the same.

(continued)

TABLE 6.4 (continued)



Poisson	Poisson(λ)
Possible applications	Number of events that occur in an interval of time when the events are occurring at a constant rate (see Sec. 6.12); number of items in a batch of random size; number of items demanded from an inventory
Mass (see Fig. 6.23)	$p(x) = \begin{cases} \frac{e^{-\lambda} \lambda^x}{x!} & \text{if } x \in \{0, 1, \dots\} \\ 0 & \text{otherwise} \end{cases}$
Distribution	$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ e^{-\lambda} \sum_{i=0}^{\lfloor x \rfloor} \frac{\lambda^i}{i!} & \text{if } x \geq 0 \end{cases}$
Parameter	$\lambda > 0$
Range	$\{0, 1, \dots\}$
Mean	λ
Variance	λ

TABLE 6.4 (continued)

Poisson	Poisson(λ)
Mode	$\begin{cases} \lambda - 1 \text{ and } \lambda & \text{if } \lambda \text{ is an integer} \\ \lfloor \lambda \rfloor & \text{otherwise} \end{cases}$
MLE	$\hat{\lambda} = \bar{X}(n)$.
Comments	- Let $Y_1, Y_2, \dots$ be a sequence of nonnegative IID random variables, and let $X = \max\{i: \sum_{j=1}^i Y_j \leq 1\}$. Then the distribution of the Y_i's is $\text{expo}(1/\lambda)$ if and only if $X \sim \text{Poisson}(\lambda)$. Also, if $X' = \max\{i: \sum_{j=1}^i Y_j \leq \lambda\}$, then the Y_i's are $\text{expo}(1)$ if and only if $X' \sim \text{Poisson}(\lambda)$ (see also Sec. 6.12). - If $X_1, X_2, \dots, X_m$ are independent random variables and $X_i \sim \text{Poisson}(\lambda_i)$, then $X_1 + X_2 + \dots + X_m \sim \text{Poisson}(\lambda_1 + \lambda_2 + \dots + \lambda_m)$.

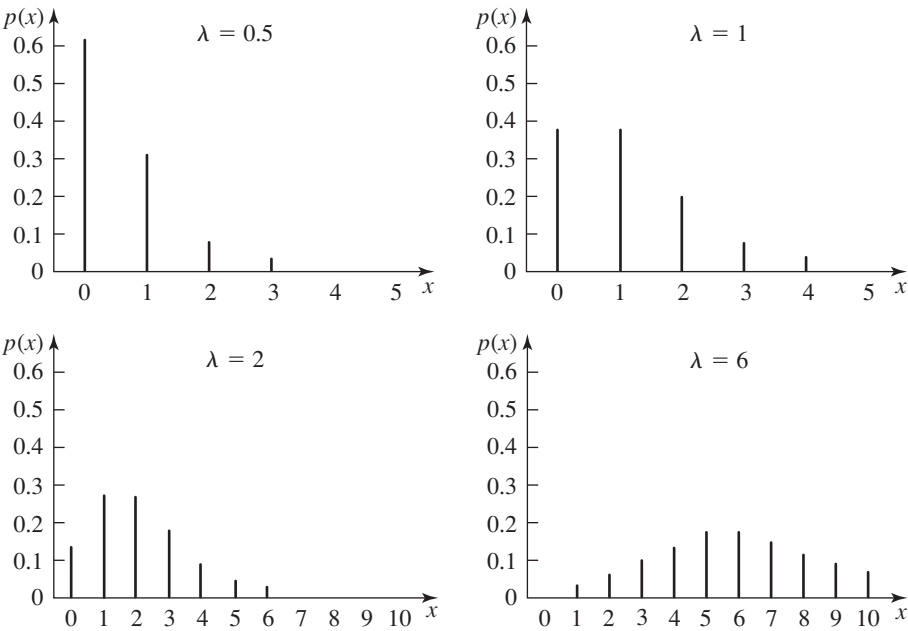
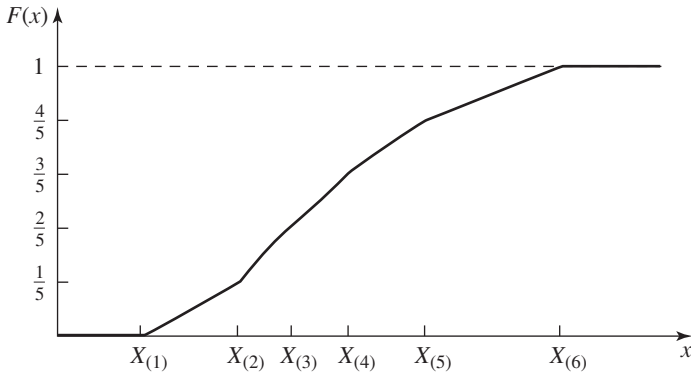


FIGURE 6.23
Poisson(λ) mass functions.

order. Let $X_{(i)}$ denote the i th smallest of the X_j 's, so that $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$. Then F is given by

$$F(x) = \begin{cases} 0 & \text{if } x < X_{(1)} \\ \frac{i-1}{n-1} + \frac{x - X_{(i)}}{(n-1)(X_{(i+1)} - X_{(i)})} & \text{if } X_{(i)} \leq x < X_{(i+1)} \\ 1 & \text{if } X_{(n)} \leq x \end{cases} \quad \text{for } i = 1, 2, \dots, n-1$$

Figure 6.24 gives an illustration for $n = 6$. Note that $F(x)$ rises most rapidly over those ranges of x in which the X_i 's are most densely distributed, as desired.

**FIGURE 6.24**

Continuous, piecewise-linear empirical distribution function from original data.

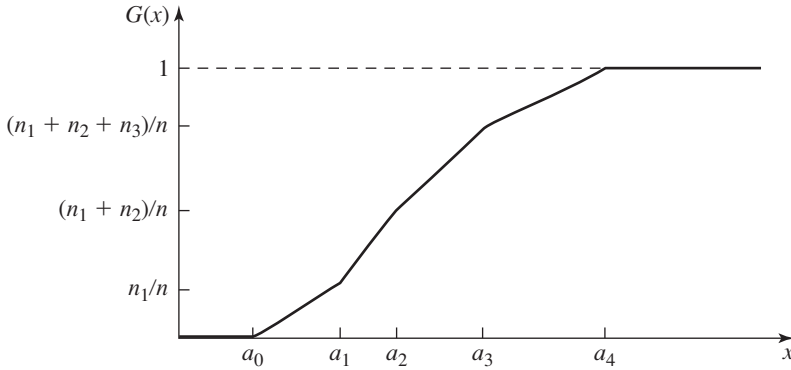
Also, for each i , $F(X_{(i)}) = (i - 1)/(n - 1)$, which is approximately (for large n) the proportion of the X_j 's that are *less than* $X_{(i)}$; this is also the way we would like a *continuous* distribution function to behave. (See Prob. 6.5 for a discussion of another way to define F .) However, one clear disadvantage of specifying this particular empirical distribution is that random values generated from it during a simulation run can never be less than $X_{(1)}$ or greater than $X_{(n)}$ (see Sec. 8.3.16). Also, the mean of $F(x)$ is not equal to the sample mean $\bar{X}(n)$ of the X_i 's (see Prob. 6.6).

If, however, the data are grouped, then a different approach must be taken since we do not know the values of the individual X_i 's. Suppose that the n X_i 's are grouped into k adjacent intervals $[a_0, a_1)$, $[a_1, a_2)$, $\dots$, $[a_{k-1}, a_k)$, so that the j th interval contains n_j observations, where $n_1 + n_2 + \dots + n_k = n$. (Often the a_j 's will be equally spaced, but we need not make this assumption.) A reasonable piecewise-linear empirical distribution function G could be specified by first letting $G(a_0) = 0$ and $G(a_j) = (n_1 + n_2 + \dots + n_j)/n$ for $j = 1, 2, \dots, k$. Then, interpolating linearly between the a_j 's, we define

$$G(x) = \begin{cases} 0 & \text{if } x < a_0 \\ G(a_{j-1}) + \frac{x - a_{j-1}}{a_j - a_{j-1}} [G(a_j) - G(a_{j-1})] & \text{if } a_{j-1} \leq x < a_j \\ 1 & \text{if } a_k \leq x \end{cases} \quad \text{for } j = 1, 2, \dots, k$$

Figure 6.25 illustrates this specification of an empirical distribution for $k = 4$. In this case, $G(a_j)$ is the proportion of the X_i 's that are less than a_j , and $G(x)$ rises most rapidly over ranges of x where the observations are most dense. The random values generated from this distribution, however, will still be bounded both below (by a_0) and above (by a_k); see Sec. 8.3.16.

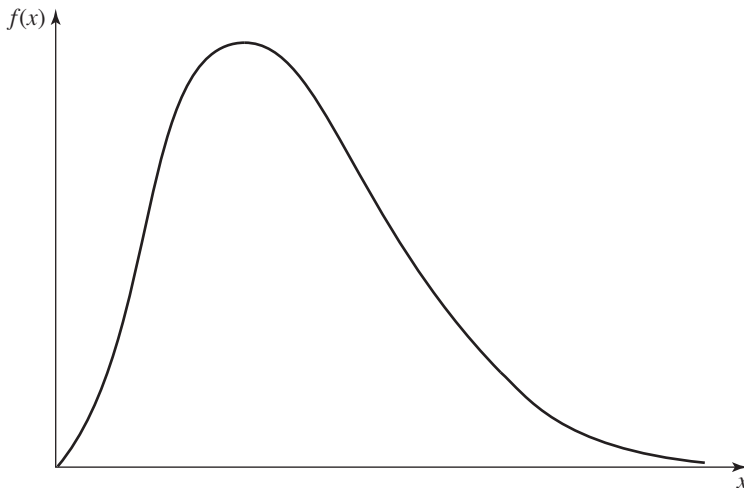
In practice, many continuous distributions are skewed to the right and have a density with a shape similar to that in Fig. 6.26. Thus, if the sample size n is not very large, we are likely to have few, if any, observations from the right tail of the true underlying distribution (since these tail probabilities are usually small). Moreover, the above empirical distributions do not allow random values to be generated

**FIGURE 6.25**

Continuous, piecewise-linear empirical distribution function from grouped data.

beyond the largest observation. On the other hand, very large generated values can have a significant impact on the disposition of a simulation run. For example, a large service time can cause considerable congestion in a queueing-type system. As a result, Bratley, Fox, and Schrage (1987, pp. 131–133, 150–151) suggest appending an exponential distribution to the right side of the empirical distribution, which allows values larger than $X_{(n)}$ to be generated.

For discrete data, it is quite simple to define an empirical distribution, provided that the original data values $X_1, X_2, \dots, X_n$ are available. For each possible value x , an empirical mass function $p(x)$ can be defined to be the proportion of the X_i 's that are equal to x . For grouped discrete data we could define a mass function such that the *sum* of the $p(x)$'s over all possible values of x in an interval is equal to the

**FIGURE 6.26**

Typical density function experienced in practice.

proportion of the X_i 's in that interval. How the individual $p(x)$'s are allocated for the possible values of x within an interval is essentially arbitrary.

6.3 TECHNIQUES FOR ASSESSING SAMPLE INDEPENDENCE

An important assumption made by many of the statistical techniques discussed in this chapter is that the observations $X_1, X_2, \dots, X_n$ are an *independent* (or *random*) sample from some underlying distribution. For example, maximum-likelihood estimation (see Sec. 6.5) and chi-square tests (see Sec. 6.6.2) assume independence. If the assumption of independence is not satisfied, then these statistical techniques may not be valid. However, even when the data are not independent, heuristic techniques such as histograms can still be used.

Sometimes observations collected over time are dependent. For example, suppose that $X_1, X_2, \dots$ represent hourly temperatures in a certain city starting at noon on a particular day. We would not expect these data to be independent, since hourly temperatures close together in time should be positively correlated. As a second example, consider the single-server queueing system in Sec. 1.4. Let $X_1, X_2, \dots$ be the delays in queue of the successive customers arriving to the system. If the arrival rate of customers is close to the service rate, the system will be congested and the X_i 's will be highly positively correlated (see Sec. 4.3).

We now describe two graphical techniques for informally assessing whether the data $X_1, X_2, \dots, X_n$ (listed in time order of collection) are independent. The *correlation plot* is a graph of the sample correlation $\hat{\rho}_j$ (see Sec. 4.4) for $j = 1, 2, \dots, l$ (l is a positive integer). The sample correlation $\hat{\rho}_j$ is an estimate of the true correlation ρ_j between two observations that are j observations apart in time. (Note that $-1 \leq \rho_j \leq 1$.) If the observations $X_1, X_2, \dots, X_n$ are independent, then $\rho_j = 0$ for $j = 1, 2, \dots, n - 1$. However, the $\hat{\rho}_j$'s will not be exactly zero even when the X_i 's are independent, since $\hat{\rho}_j$ is an observation of a random variable whose mean is not equal to 0 (see Sec. 4.4). If the $\hat{\rho}_j$'s differ from 0 by a significant amount, then this is strong evidence that the X_i 's are not independent.

The *scatter diagram* of the observations $X_1, X_2, \dots, X_n$ is a plot of the pairs (X_i, X_{i+1}) for $i = 1, 2, \dots, n - 1$. Suppose for simplicity that the X_i 's are nonnegative. If the X_i 's are independent, one would expect the points (X_i, X_{i+1}) to be scattered randomly throughout the first quadrant of the (X_i, X_{i+1}) plane. The nature of the scattering will, however, depend on the underlying distributions of the X_i 's. If the X_i 's are positively correlated, then the points will tend to lie along a line with positive slope in the first quadrant. If the X_i 's are negatively correlated, then the points will tend to lie along a line with negative slope in the first quadrant.

EXAMPLE 6.2. In Figs. 6.27 and 6.28 we give the correlation plot and scatter diagram for 100 *independent* observations from an exponential distribution with a mean of 1. Note in Fig. 6.27 that the sample correlations are close to 0, but have absolute values as large as 0.16. The scattering of the points in Fig. 6.28 substantiates the independence of the exponential data.

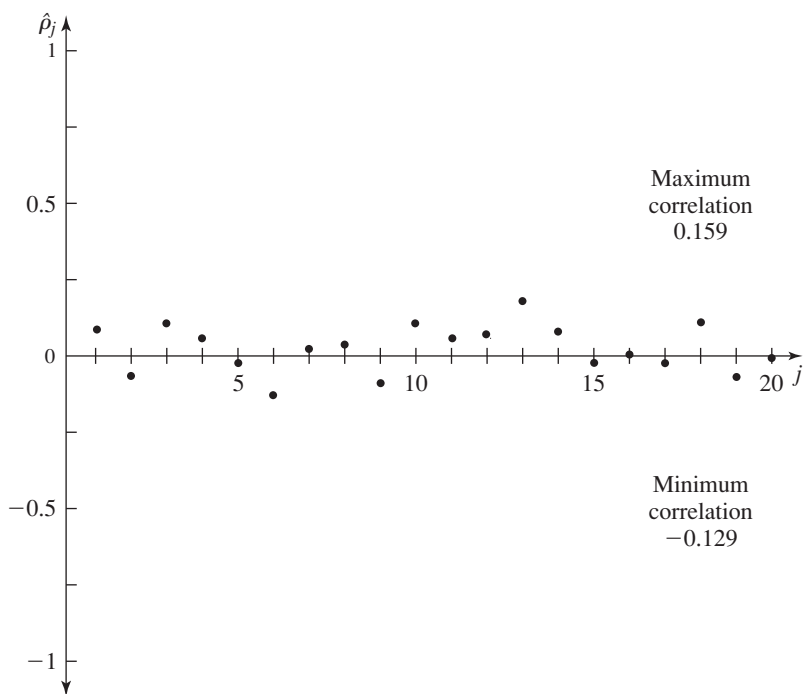


FIGURE 6.27
Correlation plot for independent exponential data.

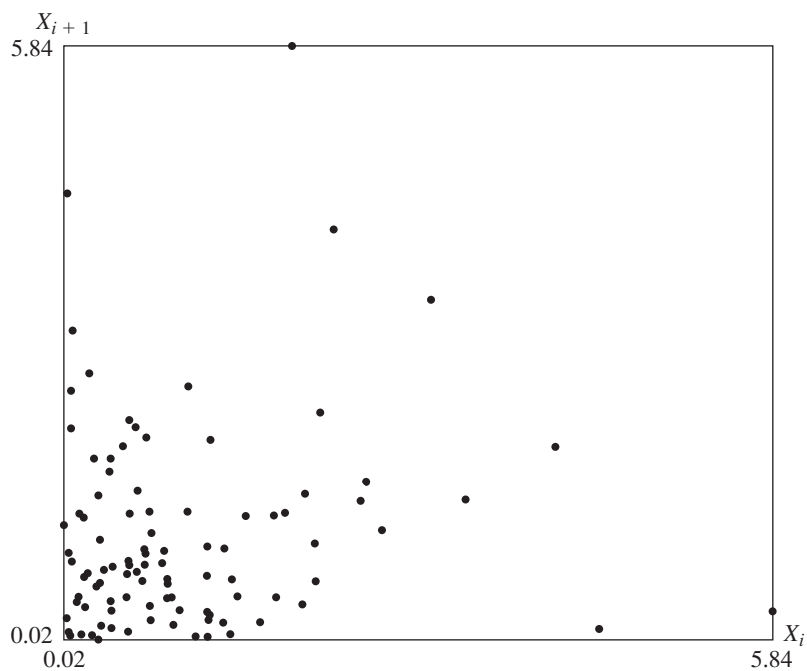
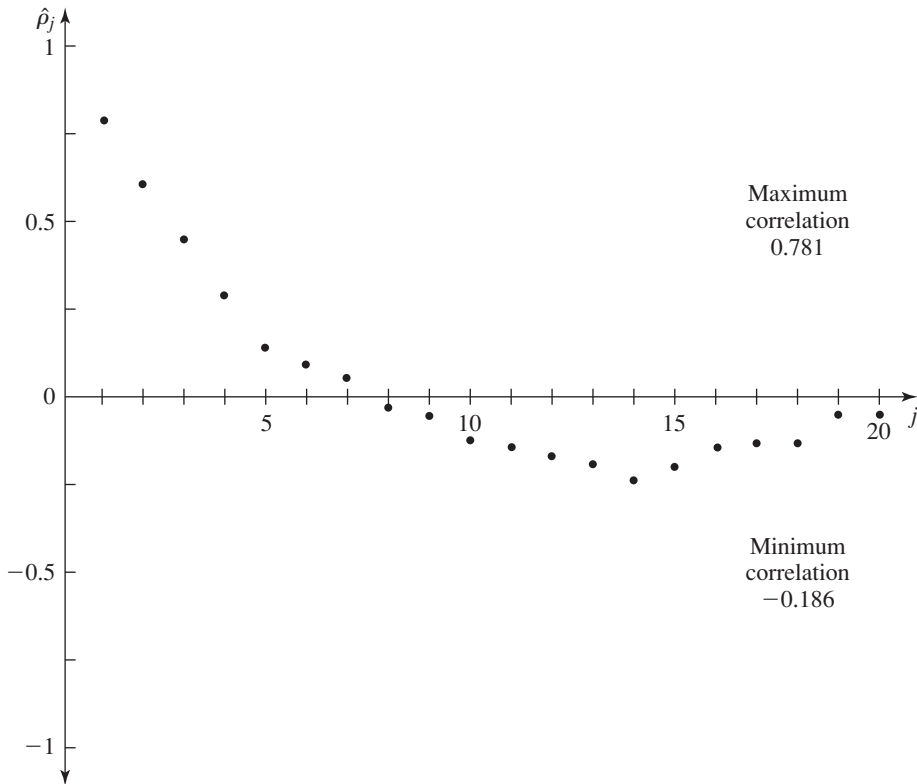


FIGURE 6.28
Scatter diagram for independent exponential data.

**FIGURE 6.29**

Correlation plot for correlated queueing data.

EXAMPLE 6.3. In Figs. 6.29 and 6.30 we present the correlation plot and scatter diagram for 100 delays in queue from an $M/M/1$ queueing system (see Sec. 1.4.3) with utilization factor $\rho = 0.8$. Note that the $\hat{\rho}_j$'s are large for small values of j and that the points in the scatter diagram tend to lie along a line with positive slope. These facts are consistent with our statement that delays in queue are positively correlated.

There are also several nonparametric (i.e., no assumptions are made about the distributions of the X_i 's) statistical tests that can be used to test formally whether $X_1, X_2, \dots, X_n$ are independent. Bartels (1982) proposes a rank version of von Neumann's ratio as a test statistic for independence and provides the necessary critical values to carry out the test. However, one potential drawback is that the test assumes that there are no "ties" in the data, where a tie means $X_i = X_j$ for $i \neq j$. This requirement will generally not be met for discrete data, and may not even be satisfied for continuous data if they are recorded with only a few decimal places of accuracy. (See the interarrival times in Table 6.7.) Bartels states that his critical values may still be reasonably accurate if the number of ties is small.

There are several versions of the runs test [see, for example, Sec. 7.4.1 and Gibbons (1985)], which can also be used to assess the independence of the X_i 's.

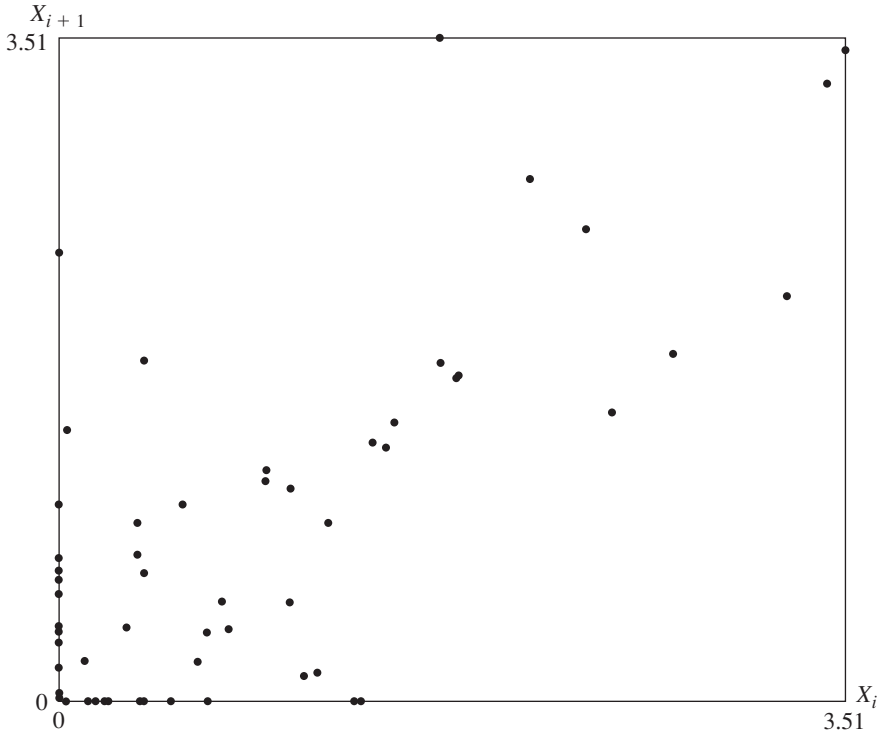


FIGURE 6.30
Scatter diagram for correlated queueing data.

They should have less difficulty with ties than the rank von Neumann test, since runs tests require only that $X_i \neq X_{i+1}$ for $i = 1, 2, \dots, n - 1$. On the other hand, Bartels showed empirically that the rank von Neumann test is considerably more powerful than one of the runs tests against certain types of alternatives to the X_i 's being independent.

6.4

ACTIVITY I: HYPOTHESIZING FAMILIES OF DISTRIBUTIONS

The first step in selecting a particular input distribution is to decide what general families—e.g., exponential, normal, or Poisson—appear to be appropriate on the basis of their shapes, without worrying (yet) about the specific parameter values for these families. This section describes some general techniques that can be used to *hypothesize* families of distributions that might be representative of a simulation input random variable.

In some situations, use can be made of *prior knowledge* about a certain random variable's role in a system to select a modeling distribution or at least rule out some

distributions; this is done on theoretical grounds and does not require any data at all. For example, if we feel that customers arrive to a service facility one at a time, at a constant rate, and so that the numbers of customers arriving in disjoint time intervals are independent, then there are theoretical reasons (see Sec. 6.12.1) for postulating that the interarrival times are IID exponential random variables. Recall also that several discrete distributions—binomial, geometric, and negative binomial—were developed from a physical model. Often the range of a distribution rules it out as a modeling distribution. Service times, for example, should not be generated directly from a normal distribution (at least in principle), since a random value from *any* normal distribution can be negative. The proportion of defective items in a large batch should not be assumed to have a gamma distribution, since proportions must be between 0 and 1, whereas gamma random variables have no upper bound. Prior information should be used whenever available, but confirming the postulated distribution with data is also strongly recommended.

In practice, we seldom have enough of this kind of theoretical prior information to select a single distribution, and the task of hypothesizing a distribution family from observed data is somewhat less structured. In the remainder of this section, we discuss various *heuristics*, or guidelines, that can be used to help one choose appropriate families of distributions.

6.4.1 Summary Statistics

Some distributions are characterized at least partially by functions of their *true* parameters. In Table 6.5 we give a number of such functions, formulas to estimate these functions from IID data $X_1, X_2, \dots, X_n$ [these estimates are called *summary* (or *descriptive*) statistics], an indication of whether they are applicable to continuous (C) or discrete (D) data, and comments about their interpretation or use. (We have included the sample minimum and maximum because of their utility, even though they may not be a direct function of a distribution's parameters.) Further discussion of many of these functions may be found in Chap. 4.

These functions might be used in some cases to suggest an appropriate distribution family. For a symmetric continuous distribution (e.g., normal), the mean μ is equal to the median $x_{0.5}$. (For a symmetric discrete distribution, the population mean and median may be only approximately equal; see the definition of the median in Sec. 4.2.) Thus, if the estimates $\bar{X}(n)$ and $\hat{x}_{0.5}(n)$ are “almost equal,” this is some indication that the underlying distribution may be symmetric. One should keep in mind that $\bar{X}(n)$ and $\hat{x}_{0.5}(n)$ are observations of random variables, and thus their relationship does not necessarily provide definitive information about the true relationship between μ and $x_{0.5}$.

The *coefficient of variation* cv can sometimes provide useful information about the form of a continuous distribution. In particular, $cv = 1$ for the exponential distribution, regardless of the scale parameter β . Thus, $\widehat{cv}(n)$ being close to 1 suggests that the underlying distribution is exponential. For the gamma and Weibull distributions, cv is greater than, equal to, or less than 1 when the shape parameter α is less than, equal to, or greater than 1, respectively. Furthermore, these distributions

TABLE 6.5
Useful summary statistics

Function	Sample estimate (summary statistic)	Continuous (C) or discrete (D)	Comments
Minimum, maximum	$X_{(1)}, X_{(n)}$	C, D	$[X_{(1)}, X_{(n)}]$ is a rough estimate of the range
Mean μ	$\bar{X}(n)$	C, D	Measure of central tendency
Median $x_{0.5}$	$\hat{x}_{0.5}(n) = \begin{cases} X_{((n+1)/2)} & \text{if } n \text{ is odd} \\ [X_{(n/2)} + X_{((n/2)+1)}]/2 & \text{if } n \text{ is even} \end{cases}$	C, D	Alternative measure of central tendency
Variance σ^2	$S^2(n)$	C, D	Measure of variability
Coefficient of variation, $cv = \frac{\sqrt{\sigma^2}}{\mu}$	$\hat{cv}(n) = \frac{\sqrt{S^2(n)}}{\bar{X}(n)}$	C	Alternative measure of variability
Lexis ratio, $\tau = \frac{\sigma^2}{\mu}$	$\hat{\tau}(n) = \frac{S^2(n)}{\bar{X}(n)}$	D	Alternative measure of variability
Skewness, $\nu = \frac{E[(X - \mu)^3]}{(\sigma^2)^{3/2}}$	$\hat{\nu}(n) = \frac{n^2}{(n-1)(n-2)} \frac{\sum_{i=1}^n [X_i - \bar{X}(n)]^3/n}{[S^2(n)]^{3/2}}$	C, D	Measure of symmetry

will have a shape similar to the density function in Fig. 6.26 when $\alpha > 1$, which implies that $cv < 1$. On the other hand, the lognormal distribution always has a density with a shape similar to that in Fig. 6.26, but its cv can be any positive real number. Thus, if the underlying distribution (observed histogram) has this shape and $\hat{cv}(n) > 1$, the lognormal may be a better model than either the gamma or Weibull. For the remainder of the distributions in Table 6.3, the cv is not particularly useful. [In fact, cv is not even well defined for distributions such as $U(-c, c)$ (for $c > 0$) or $N(0, \sigma^2)$, since the mean μ is zero.]

For a discrete distribution, the *lexis ratio* τ plays the same role that the coefficient of variation does for a continuous distribution. We have found the lexis ratio to be very useful in discriminating among the Poisson, binomial, and negative binomial distributions, since $\tau = 1$, $\tau < 1$, and $\tau > 1$, respectively, for these distributions. (Note that the geometric distribution is a special case of the negative binomial.)

The *skewness* ν is a measure of the symmetry of a distribution. For symmetric distributions like the normal, $\nu = 0$. If $\nu > 0$ (e.g., $\nu = 2$ for the exponential distribution), the distribution is skewed to the right; if $\nu < 0$, the distribution is skewed to the left. Thus, the estimated skewness $\hat{\nu}(n)$ [see Joanes and Gill (1998)] can be used to ascertain the shape of the underlying distribution. Our experience indicates that many distributions encountered in practice are skewed to the right.

It is possible to define another function of a distribution's parameters, called the *kurtosis*, which is a measure of the "tail weight" of a distribution [see, for example, Kendall, Stuart, and Ord (1987, pp. 107–108)]. However, we have not found the kurtosis to be very useful for discriminating among distributions.

6.4.2 Histograms

For a continuous data set, a histogram is essentially a graphical estimate (see the discussion below) of the plot of the density function corresponding to the distribution of our data $X_1, X_2, \dots, X_n$. Density functions, as shown in Figs. 6.5 through 6.16, tend to have recognizable shapes in many cases. Therefore, a graphical estimate of a density should provide a good clue to the distributions that might be tried as a model for the data.

To make a *histogram*, we break up the range of values covered by the data into k disjoint adjacent intervals $[b_0, b_1), [b_1, b_2), \dots, [b_{k-1}, b_k)$. All the intervals should be the same width, say, $\Delta b = b_j - b_{j-1}$, which might necessitate throwing out a few extremely large or small X_i 's to avoid getting an unwieldy-looking histogram plot. For $j = 1, 2, \dots, k$, let h_j be the *proportion* of the X_i 's that are in the j th interval $[b_{j-1}, b_j)$. Finally, we define the function

$$h(x) = \begin{cases} 0 & \text{if } x < b_0 \\ h_j & \text{if } b_{j-1} \leq x < b_j \\ 0 & \text{if } b_k \leq x \end{cases} \quad \text{for } j = 1, 2, \dots, k$$

which we plot as a function of x . (See Example 6.4 below for an illustration of a histogram.) The plot of h , which is piecewise-constant, is then compared with plots of densities of various distributions on the basis of shape alone (location and scale

differences are ignored) to see what distributions have densities that resemble the histogram h .

To see why the shape of h should resemble the true density f of the data, let X be a random variable with density f , so that X is distributed as the X_i 's. Then, for any fixed j ($j = 1, 2, \dots, k$),

$$P(b_{j-1} \leq X < b_j) = \int_{b_{j-1}}^{b_j} f(x) dx = \Delta b f(y)$$

for some number $y \in (b_{j-1}, b_j)$. (The first equation is by the definition of a continuous random variable, and the second follows from the mean-value theorem of calculus.) On the other hand, the probability that X falls in the j th interval is approximated by h_j , which is the value of $h(y)$. Therefore,

$$h(y) = h_j \approx \Delta b f(y)$$

so that $h(y)$ is roughly proportional to $f(y)$; that is, h and f have roughly the same shape. (Actually, an *estimate* of f is obtained by dividing the function h by the constant Δb .)

Histograms are applicable to any continuous distribution and provide a readily interpreted visual synopsis of the data. Furthermore, it is easy to “eyeball” a histogram in reference to certain density functions. There are, however, certain difficulties. Most vexing is the absence of a definitive guide for choosing the number of intervals k (or, equivalently, their width Δb).

Several rules of thumb have been suggested for choosing the number of intervals k {e.g., Sturges's rule [see Hoaglin, Mosteller, and Tukey (1983, pp. 23–24)] and a normal approximation due to Scott (1979)}. The best known of these guidelines is probably Sturges's rule, which says that k should be chosen according to the following formula:

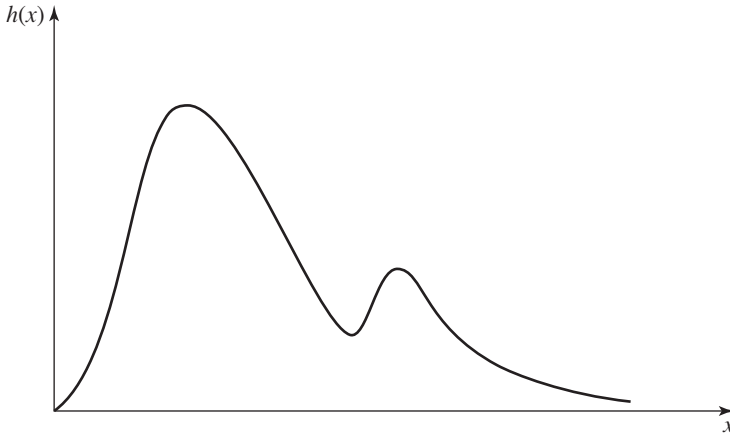
$$k = \lfloor 1 + \log_2 n \rfloor = \lfloor 1 + 3.322 \log_{10} n \rfloor$$

However, in general, we do not believe that such rules are very useful (see Example 6.4). We recommend trying several different values of Δb and choosing the smallest one that gives a “smooth” histogram. This is clearly a matter of some subjectivity and represents the major problem in using histograms. If Δb is chosen too small, the histogram will have a “ragged” shape since the variances of the h_j 's will be large. If Δb is chosen too large, then the histogram will have a “block-like” shape, and the true shape of the underlying density will be masked since the data have been overaggregated. In particular, a large spike in the density function near $x = 0$ or elsewhere (see Fig. 6.12) could be missed if Δb is too large.

As we have noted, a histogram is an estimate (except for rescaling) of the density function. There are many other ways in which the density function can be estimated from data, some of which are quite sophisticated. We refer the interested reader to the survey paper of Wegman (1982, pp. 309–315) and the book by Silverman (1986).

The probability mass function corresponding to a discrete data set can also be estimated by using a histogram. For each possible value x_j that can be assumed by the data, let h_j be the proportion of the X_i 's that are equal to x_j . Vertical bars of height h_j are plotted versus x_j , and this is compared with the mass functions of the discrete distributions in Sec. 6.2.3 on the basis of shape.

For a discrete data set, h_j (which is a random variable) is an unbiased estimator of $p(x_j)$, where $p(x)$ is the true (unknown) mass function of the data (see Prob. 6.7).

**FIGURE 6.31**

Histogram corresponding to a density function with two local modes.

However, in this case we need not make any arbitrary subjective decisions about interval width and placement.

There are certain situations in which the histogram will have several local modes (or “humps”), in which case none of the standard distributions discussed in Sec. 6.2 may provide an adequate representation. In Fig. 6.31 we give such an example, which might represent (continuous) times to repair some machine collected over a 1-year period. There are two types of breakdowns for the machine. Most of the breakdowns are minor, and the corresponding repair time is relatively small; this case corresponds to the left hump in Fig. 6.31. A small proportion of the breakdowns are major and have large repair times, since spare parts need to be ordered. This results in the right hump in Fig. 6.31. If the observed repair times can be separated into these two cases (corresponding to minor and major repairs), with p_j being the proportion of observations for case j ($j = 1, 2$), then a density $f_j(x)$ is fit to the class j observations ($j = 1, 2$) using the methods discussed in Secs. 6.4 through 6.6. Thus, the overall repair-time density $f(x)$ is given by

$$f(x) = p_1 f_1(x) + p_2 f_2(x)$$

and random repair times can be generated from $f(x)$ during a simulation by using the composition technique (see Sec. 8.2.2). See Sec. 6.9 and Prob. 6.8 for alternative methods for modeling a histogram with several modes.

6.4.3 Quantile Summaries and Box Plots

The *quantile summary* [see, for example, Tukey (1970)] is a synopsis of the sample that is useful for determining whether the underlying probability density function or probability mass function is symmetric or skewed to the right or to the left. It is applicable to either continuous or discrete data sets; however, for expository convenience we will explain it only for the continuous case.

TABLE 6.6

Structure of the quantile summary for the sample $X_1, X_2, \dots, X_n$

Quantile	Depth	Sample value(s)		Midpoint
Median	$i = (n + 1)/2$	$X_{(i)}$		$X_{(i)}$
Quartiles	$j = (\lfloor i \rfloor + 1)/2$	$X_{(j)}$	$X_{(n-j+1)}$	$[X_{(j)} + X_{(n-j+1)}]/2$
Octiles	$k = (\lfloor j \rfloor + 1)/2$	$X_{(k)}$	$X_{(n-k+1)}$	$[X_{(k)} + X_{(n-k+1)}]/2$
Extremes	1	$X_{(1)}$	$X_{(n)}$	$[X_{(1)} + X_{(n)}]/2$

Suppose that $F(x)$ is the distribution function for a continuous random variable. Suppose further that $F(x)$ is continuous and strictly increasing when $0 < F(x) < 1$. [This means that if $x_1 < x_2$ and $0 < F(x_1) \leq F(x_2) < 1$, then in fact $F(x_1) < F(x_2)$.] For $0 < q < 1$, the q -quantile of $F(x)$ is that number x_q such that $F(x_q) = q$. If F^{-1} denotes the inverse of $F(x)$, then $x_q = F^{-1}(q)$. $\{F^{-1}$ is that function such that $F[F^{-1}(x)] = F^{-1}[F(x)] = x\}$. Here we are particularly interested in the median $x_{0.5}$, the lower and upper quartiles $x_{0.25}$ and $x_{0.75}$, and the lower and upper octiles $x_{0.125}$ and $x_{0.875}$. A quantile summary for the sample $X_1, X_2, \dots, X_n$ has a form that is given in Table 6.6.

In Table 6.6, if the “depth” subscript l (l is equal to i, j , or k) is halfway between the integers m and $m' = m + 1$, then $X_{(l)}$ is defined to be the average of $X_{(m)}$ and $X_{(m')}$. The value $X_{(i)}$ is an estimate of the median, $X_{(j)}$ and $X_{(n-j+1)}$ are estimates of the quartiles, and $X_{(k)}$ and $X_{(n-k+1)}$ are estimates of the octiles. If the underlying distribution of the X_i 's is symmetric, then the four midpoints should be approximately equal. On the other hand, if the underlying distribution of the X_i 's is skewed to the right (left), then the four midpoints (from the top to the bottom of the table) should be increasing (decreasing).

A *box plot* is a graphical representation of the quantile summary (see Fig. 6.33). Fifty percent of the observations fall within the horizontal boundaries of the box $[x_{0.25}, x_{0.75}]$. (In some books, box plots do not contain octiles.)

The following two examples illustrate the use of the techniques discussed in Sec. 6.4.

EXAMPLE 6.4. A simulation model was developed for a drive-up banking facility, and data were collected on the arrival pattern for cars. Over a fixed 90-minute interval, 220 cars arrived, and we noted the (continuous) interarrival time X_i (in minutes) between cars i and $i + 1$, for $i = 1, 2, \dots, 219$. Table 6.7 lists these $n = 219$ interarrival times after they have been sorted into increasing order. The numbers of cars arriving in the six consecutive 15-minute intervals were counted and found to be approximately equal, suggesting that the arrival rate is somewhat constant over *this* 90-minute interval. Furthermore, cars arrive one at a time, and there is no reason to believe that the numbers of arrivals in disjoint intervals are not independent. Thus, on theoretical grounds (see Sec. 6.12.1) we postulate that the interarrival times are exponential. To substantiate this hypothesis, we first look at the summary statistics given in Table 6.8. Since $\bar{X}(219) = 0.399 > 0.270 = \hat{x}_{0.5}(219)$ and $\hat{\nu}(219) = 1.478$, this suggests that the underlying distribution is skewed to the right, rather than symmetric. Furthermore, $\hat{c}(219) = 0.953$, which is close to the theoretical value of 1 for the exponential distribution. Next we made three different histograms of the data, using $b_0 = 0$ in each case and $\Delta b = 0.050, 0.075$, and 0.100 , as shown in Fig. 6.32 (see also Fig. 6.2). The smoothest-looking

TABLE 6.7
n = 219 interarrival times (minutes) sorted into increasing order

0.01	0.06	0.12	0.23	0.38	0.53	0.88
0.01	0.07	0.12	0.23	0.38	0.53	0.88
0.01	0.07	0.12	0.24	0.38	0.54	0.90
0.01	0.07	0.13	0.25	0.39	0.54	0.93
0.01	0.07	0.13	0.25	0.40	0.55	0.93
0.01	0.07	0.14	0.25	0.40	0.55	0.95
0.01	0.07	0.14	0.25	0.41	0.56	0.97
0.01	0.07	0.14	0.25	0.41	0.57	1.03
0.02	0.07	0.14	0.26	0.43	0.57	1.05
0.02	0.07	0.15	0.26	0.43	0.60	1.05
0.03	0.07	0.15	0.26	0.43	0.61	1.06
0.03	0.08	0.15	0.26	0.44	0.61	1.09
0.03	0.08	0.15	0.26	0.45	0.63	1.10
0.04	0.08	0.15	0.27	0.45	0.63	1.11
0.04	0.08	0.15	0.28	0.46	0.64	1.12
0.04	0.09	0.17	0.28	0.47	0.65	1.17
0.04	0.09	0.18	0.29	0.47	0.65	1.18
0.04	0.10	0.19	0.29	0.47	0.65	1.24
0.04	0.10	0.19	0.30	0.48	0.69	1.24
0.05	0.10	0.19	0.31	0.49	0.69	1.28
0.05	0.10	0.20	0.31	0.49	0.70	1.33
0.05	0.10	0.21	0.32	0.49	0.72	1.38
0.05	0.10	0.21	0.35	0.49	0.72	1.44
0.05	0.10	0.21	0.35	0.50	0.72	1.51
0.05	0.10	0.21	0.35	0.50	0.74	1.72
0.05	0.10	0.21	0.36	0.50	0.75	1.83
0.05	0.11	0.22	0.36	0.51	0.76	1.96
0.05	0.11	0.22	0.36	0.51	0.77	
0.05	0.11	0.22	0.37	0.51	0.79	
0.06	0.11	0.23	0.37	0.52	0.84	
0.06	0.11	0.23	0.38	0.52	0.86	
0.06	0.12	0.23	0.38	0.53	0.87	

TABLE 6.8
Summary statistics for the
interarrival-time data

Summary statistic	Value
Minimum	0.010
Maximum	1.960
Mean	0.399
Median	0.270
Variance	0.144
Coefficient of variation	0.953
Skewness	1.478

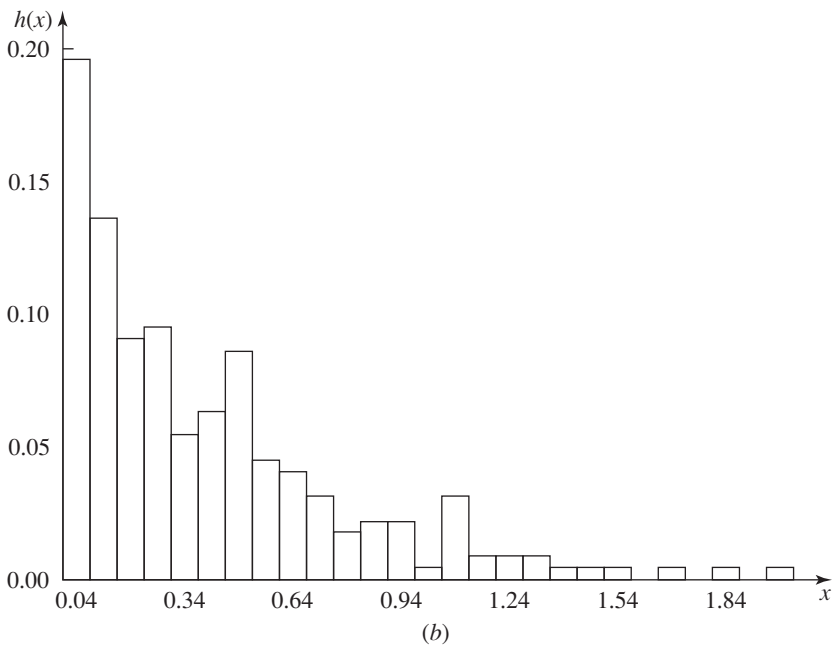
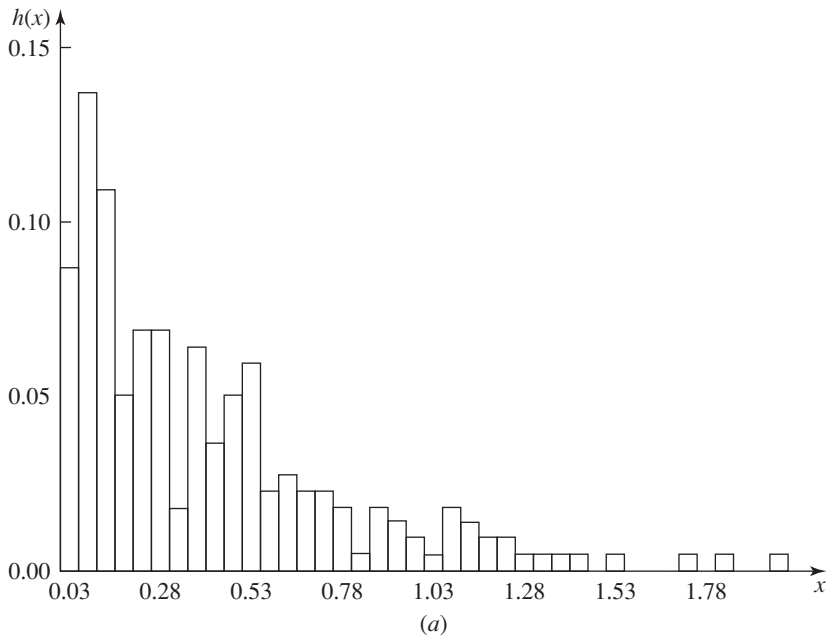


FIGURE 6.32
Histograms of the interarrival-time data in Table 6.7: (a) $\Delta b = 0.050$;
(b) $\Delta b = 0.075$; (c) $\Delta b = 0.100$.

(continued)

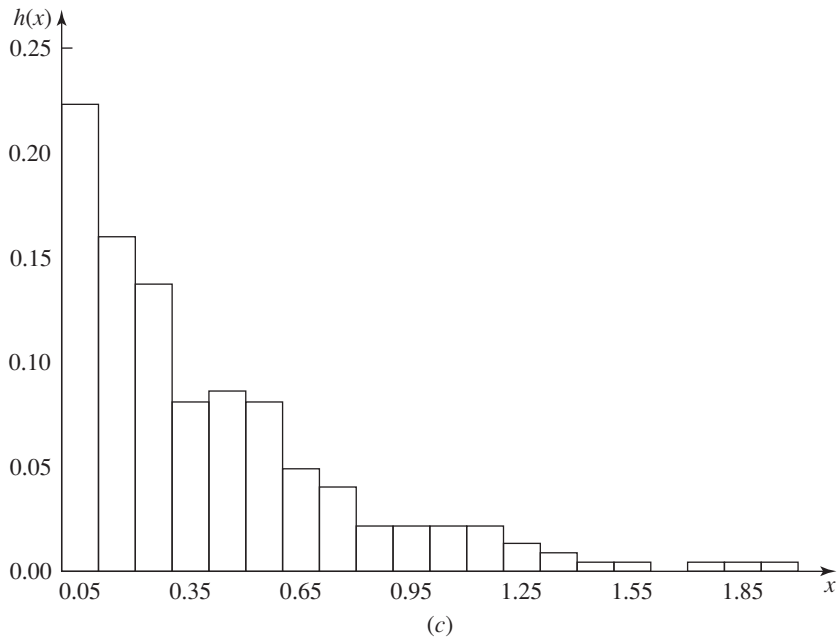


FIGURE 6.32
(continued)

histogram appears to be for $\Delta b = 0.100$ and its shape resembles that of an exponential density. (Note that Sturges's rule gives $k = 8$ and $\Delta b = 0.250$, resulting in overaggregation of the data.) Finally, in Fig. 6.33 we give the quantile summary and box plot for the interarrival times. The increasing midpoints and the elongated nature of the right side of the box plot reaffirm that the underlying distribution is exponential. In summary, for both theoretical and empirical reasons, we hypothesize that the interarrival times are exponential.

Quantile	Depth	Sample value(s)		Midpoint
Median	110	0.270		0.270
Quartiles	55.5	0.100	0.545	0.323
Octiles	28	0.050	0.870	0.460
Extremes	1	0.010	1.960	0.985

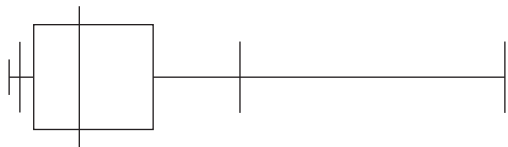


FIGURE 6.33
Quantile summary and box plot for the interarrival-time data.

TABLE 6.9
Values and counts for $n = 156$ demand sizes arranged into increasing order

0(59),	1(26),	2(24),	3(18),	4(12),
5(5),	6(4),	7(3),	9(3),	11(2)

TABLE 6.10
Summary statistics for the demand-size data

Summary statistic	Value
Minimum	0.000
Maximum	11.000
Mean	1.891
Median	1.000
Variance	5.285
Lexis ratio	2.795
Skewness	1.687

EXAMPLE 6.5. Table 6.9 gives the values and counts for $n = 156$ observations on the (discrete) number of items demanded in a week from an inventory over a 3-year period, arranged into increasing order. Rather than giving all the individual values, we give the frequency counts; 59 X_i 's were equal to 0, 26 X_i 's were equal to 1, etc. Summary statistics and a histogram for these data are given in Table 6.10 and Fig. 6.34, respectively.

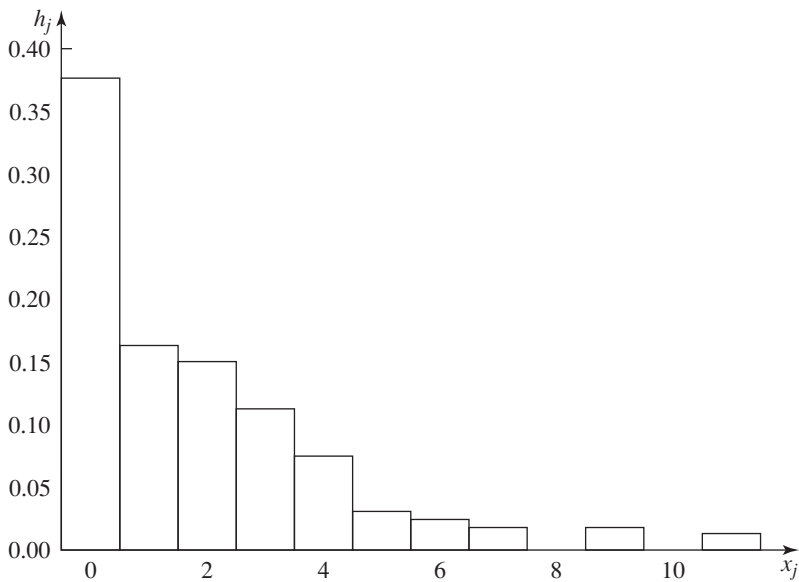


FIGURE 6.34
Histogram of the demand-size data in Table 6.9.