Chapter 3

Elements of Probability and Statistics

Many real-life systems exhibit behavior with random elements. Such systems encompass a vast array of application areas, such as the following:

1. Manufacturing
   - Random demand for product held in an inventory system
   - Random product processing time or transfer time
   - Random machine failures and repairs
2. Transportation
   - Random congestion on a highway
   - Random weather patterns
   - Random travel times between pairs of origination and destination points
3. Telecommunications
   - Random traffic arriving at a telecommunications network
   - Random transmission time (depending on available resources, such as buffer space and CPU)

Indeed, simulation modeling with random elements is often referred to as *Monte Carlo simulation*, presumably after its namesake casino at Monte Carlo on the Mediterranean. This apt term commemorates the link between randomness and gambling, going back to the French scientist Blaise Pascal in the 17th century.

Formally, modeling a random system as a discrete-event simulation simply means that randomness is introduced into events in two basic ways:

- Event occurrence times may be random.
- Event state transitions may be random.

For instance, random interarrival times at a manufacturing station exemplify the first case, while random destinations of product units emerging from an inspection station (possibly needing re-work with some probability) exemplify the second. Either way, probability and statistics are fundamental to simulation models and to understanding the underlying random phenomena in a real-life system under study. In particular, they play a key role in simulation-related input analysis and output analysis. Recall that input

analysis models random components by fitting a probabilistic model to empirical data generated by the system under study, or by postulating a model when empirical data is lacking or insufficient. Once input analysis is complete and simulation runs (replications) are generated, output analysis is then employed to verify and validate the simulation model, and to generate statistical predictions for performance measures of interest.

This chapter reviews the basic probabilistic and statistical concepts underlying Monte Carlo simulation. Additional material will be presented in Chapter 7 (on input analysis), Chapter 8 (on model verification and validation), Chapter 9 (on output analysis), and Chapter 10 (on correlation analysis). For further readings on probability, see, for example, Çinlar (1975), Ross (1993), Hoel et al. (1971a), Feller (1968), and Taylor and Karlin (1984).

3.1 ELEMENTARY PROBABILITY THEORY

Informally, probability is a measure of the uncertainty inherent in the occurrence of random phenomena, such as the following statements in future tense:

- It will rain tomorrow.
- I will win the lottery next week.
- The Fed will raise interest rates next month.

Probability is measured on a continuous scale spanning the interval [0, 1]. In particular, a probability of 0 means that it is certain that the phenomenon *will not* occur, while a probability of 1 means that it is certain that the phenomenon *will* occur. Probabilities lying strictly between 0 and 1 quantify any intermediate likelihood of occurrence.

The notion of "likelihood" has a practical operational meaning, linked intimately with statistics. Suppose we observe multiple "experiments" in the underlying system (replications), and each time we record whether or not some specified phenomenon, A, occurred. Suppose we observed n such experiments and found that in k of them the phenomenon A occurred (and therefore, in $n - k$ of them, it did not occur). The probability of A occurring, is then estimated by the frequency ratio

$$\hat{p}_A = \frac{k}{n},$$

which is indeed between 0 and 1. This is merely an estimate with a likely experimental error, but we hope that as the number of experiments n increases, the accuracy of $\hat{\boldsymbol{p}}_{\boldsymbol{A}}$ would improve. In practice, people often use the term "probability" loosely to refer to its estimate, because the true probability is unknown.

Probability estimates can be more complex than simply frequency ratios. For example, a probability estimate of a horse winning a race can indeed be computed as a ratio based on the horse's historical track record. However, the odds published by book makers are estimates based on the opinion of the betting public, which is itself based on many other complex factors, such as past races, weather, trainers, and possibly illegal inside information. All these are encapsulated into a measurable quantity and an observable statistic.

3.1.1 Probability Spaces

The theory of probability is an abstraction that formalizes these ideas. It introduces a set of postulates, including a probability calculus. Formally, a probability space is a triple of objects, $(\Omega, E, \Pr)$, where:

- Ω is the *sample space*, corresponding to all possible "outcomes" of the random phenomenon under consideration. Although the sample space is an abstract concept, a *sample point* $\omega \in \Omega$ can be thought of as an "experiment" over the underlying (random) system.
- E is the *event set*, corresponding to permissible sets of "outcomes." Thus, an event $A \in E$ is a set of sample points, that is, $A \subset \Omega$. The empty set, ϕ, and the sample space, Ω, always belong to E. Furthermore, if Ω is countable (finite or infinite), then every subset of Ω belongs to E. In all other cases, we must impose technical conditions on the membership of events, which are beyond the scope of this book.
- Pr is a *probability measure*, which satisfies the following postulates:
  a. $0 \leq \Pr\{A\} \leq 1$ for all $A \in E$ (in particular, $\Pr\{\phi\} = 0$ and $\Pr\{\Omega\} = 1$).
  b. For any events $A, B \in E$, satisfying $A \cap B = \phi$ (disjoint events),

$$\Pr\{A \cup B\} = \Pr\{A\} + \Pr\{B\}, \tag{3.1}$$

  which is a special case of the equality

$$\Pr\{A \cup B\} = \Pr\{A\} + \Pr\{B\} - \Pr\{A \cap B\}. \tag{3.2}$$

The postulates above are reasonable. The probability of "no outcome" is impossible and therefore always evaluates to the minimal value, 0. The probability of any of the "possible outcomes" occurring is a certainty, and therefore always evaluates to the maximal value, 1. Finally, if two events do not overlap, their probability of occurrence is the sum of their probabilities. Otherwise, the sum of their probabilities contains twice the probability of their intersection (instead of one such probability), so one superfluous probability of the intersection is subtracted.

Let $\Omega = \cup A_n$ be a partition of the sample space into mutually disjoint events $\{A_n\}$. Then for any event B, the *formula of total probability* is

$$\Pr\{B\} = \sum_n \Pr\{A_n \cap B\}. \tag{3.3}$$

3.1.2 Conditional Probabilities

The concept of *conditioning* plays a major role in probability theory. More precisely, if A and B are events, such that $\Pr\{B\} > 0$, then the probability of event A *conditioned* on event B, is denoted by $\Pr\{A|B\}$ and defined by

$$\Pr\{A|B\} = \frac{\Pr\{A \cap B\}}{\Pr\{B\}}. \tag{3.4}$$

Equation 3.4 is alternatively referred to as the probability of event A *given* event B. The meaning of conditional probabilities can be explained as follows. Suppose we wish to consider the occurrence of event A, but only if we know that a prescribed event B has actually occurred. In a sense, we require the event B to become our new sample space,

and we look at the probability of event A only when it occurs concurrently with event B (the numerator of Eq. 3.4). We divide by the probability of B (the denominator of Eq. 3.4), to ensure that the maximal value of the conditional probability, $\Pr\{B|B\}$, is normalized to 1. Thus, a conditional probability cannot be defined via Eq. 3.4 when the given (conditioning) event, B, has zero probability of occurring; in this case, the conditional probability should be specified by other means.

The operational meaning of conditioning can be viewed as the modification of the probability of an event A by the added "information" that another event B has actually occurred. For example, if we throw a single die, then the odds of the outcome being 4 is one in six (probability 1/6). However, suppose that after the die is cast, we are not allowed to see the outcome, but are told that the outcome was even. This new information modifies the previous probability, $\Pr\{\text{outcome is } 4\} = 1/6$, to a new probability, $\Pr\{\text{outcome is } 4|\text{outcome is even}\} = 1/3$, since the odds of obtaining an even outcome (2, 4, or 6) is 1 in 3 (note that these events are disjoint, so the probabilities are additive). By the same token, if it were known that the outcome turned out to be odd, then $\Pr\{\text{outcome is } 4|\text{outcome is odd}\} = 0$. If, however, we were told that the outcome was a two-digit number (an impossible event), we would not be able to define the conditional probability.

3.1.3 Dependence and Independence

The concepts of *event independence* and *event dependence* are expressed in terms of conditional probabilities. A set of events, $A_i, i = 1, 2, \ldots, n$, are said to be (mutually) *independent*, provided that

$$\Pr\{A_1, A_2, \ldots, A_n\} = \prod_{i=1}^{n} \Pr\{A_i\}, \tag{3.5}$$

where the notation $\Pr\{A_1, A_2, \ldots, A_n\}$ is shorthand for $\Pr\{A_1 \cap A_2 \cap \ldots \cap A_n\}$. Otherwise, the events are said to be dependent. For two events, A and B, Eq. 3.5 can be written as

$$\Pr\{A \cap B\} = \Pr\{A\} \times \Pr\{B\}. \tag{3.6}$$

The meaning of independence (or dependence) becomes clearer when we divide (when permissible) both sides of the above equation by $\Pr\{A\}$, and separately by $\Pr\{B\}$. We then obtain the dual equations

$$\Pr\{A|B\} = \Pr\{A\} \text{or} \Pr\{B|A\} = \Pr\{B\}, \tag{3.7}$$

each of which is equivalent to Eq. 3.6. Thus, from Eq. 3.7, independence holds when the conditional and unconditional probabilities are equal. In other words, knowledge of one event does not modify the (unconditioned) probability of the other event. Presumably, this is so because the two events are "unrelated," and the occurrence of one does not affect the odds of the other. Conversely, two events are dependent, if knowledge of one event modifies the probability of the other.

Be careful not to confuse independent events ($\Pr\{A \cap B\} = \Pr\{A\} \times \Pr\{B\}$) with disjoint events ($\{A \cap B\} = \phi$). These are entirely different concepts, none of which necessarily implies the other.

3.2 RANDOM VARIABLES

Conducting an experiment can be thought of as sampling an observation at a sample point, subject to some underlying probability. For example, suppose we select at random a car on the assembly line for quality assurance. We can then make multiple measurements on the car, each revealing a different aspect of its quality, including possibly the following:

- Breaking distance at 65 miles per hour
- Extent of tire wear after 50,000 miles
- Crash test performance

The concept of *random variable* is the theoretical construct that captures aspects of sample points. In the simulation context, a random variable is also referred to as a *variate*. It should be pointed out that even though practitioners do not always refer explicitly to an underlying probability space, such a space is always assumed implicitly.

Omitting some technical conditions, which are beyond the scope of this book, a random variable X is a function

$$X: \Omega \to S, \tag{3.8}$$

where Ω is the underlying sample space, and S is called the *state space* of X, and consists of all possible values that X can assume. A particular value, $X(\omega) = x \in S$, realized by a random variable for a particular sample point, ω ("experiment outcome"), is called a *realization* of X. For example, a particular car in a road test plays the role of a sample point, ω, while its properties (breaking distance, tire wear, etc.) correspond to realizations of various random variables. Note carefully that the notion of a random variable is quite distinct from the notion of its realizations. To keep this distinction typographically clear, we shall always denote realizations by lower-case letters and random variables by upper-case letters.

A state space S can be quite general. It may be real valued or vector valued. In fact, it need not be numerical at all in order to capture qualitative aspects. For example, if the random variable X represents the status of a machine, the corresponding state space may be defined as the four status descriptors $S = \{\text{Idle}, \text{Busy}, \text{Down}, \text{Blocked}\}$.

Random variables are classified according to their associated state space. A state space is said to be *discrete* if it is countable, or *continuous*, if it is not (it can also be *mixed* with discrete and continuous components). For example, the status indicators $S = \{\text{Up}, \text{Down}\}$ for a machine form a discrete state space. However, the random variable that measures the time to failure of the machine has a continuous state space, since it can take values in some interval $S = [0, T_{\max}]$ of non-negative real numbers.

3.3 DISTRIBUTION FUNCTIONS

The probabilistic properties of random variables are characterized by their *distribution functions* (often abbreviated to *distributions*). These functions assume various forms, depending on the type of the associated random variable and the nature of its state space (numerical or not). In particular, a distribution function is continuous or discrete (or mixed) according to the type of its associated random variable.

3.3.1 Probability Mass Functions

Every discrete random variable X has an associated *probability mass function* (*pmf*), $p_X(x)$, defined by

$$p_X(x) = \Pr\{X = x\}, \quad x \in S. \tag{3.9}$$

Note that the notation $\{X = x\}$ above is a shorthand notation for the event $\{\omega: X(\omega) = x\}$. It should be pointed out that the technical definition of a random variable ensures that this set is actually an event (i.e., belongs to the underlying event set E). Thus, the pmf is always guaranteed to exist, and has the following properties:

$$0 \leq p_X(x) \leq 1, \quad x \in S,$$

and

$$\sum_{x \in S} p_X(x) = 1.$$

3.3.2 Cumulative Distribution Functions

Every real-valued random variable X (discrete or continuous) has an associated *cumulative distribution function* (*cdf*), $F_X(x)$, defined by

$$F_X(x) = \Pr\{X \leq x\}, \ -\infty < x < \infty. \tag{3.10}$$

Note that the notation $\{X \leq x\}$ is a shorthand notation for the event $\{\omega: X(\omega) \leq x\}$. It should be pointed out that the technical definition of a random variable ensures that this set is actually an event (i.e., belongs to the underlying event set E). Thus, the cdf is always guaranteed to exist.

The cdf has the following properties:

(i) $0 \leq F_X(x) \leq 1, \ -\infty < x < \infty$.
(ii) $\lim_{x \to -\infty} F_X(x) = 0$ and $\lim_{x \to \infty} F_X(x) = 1$.
(iii) If $x_1 \leq x_2$, then $F_X(x_1) \leq F_X(x_2)$ (monotonicity).

Since $\{X \leq x_1\}$ is contained in $\{X \leq x_2\}$, this implies the formula

$$\Pr\{x_1 \leq X \leq x_2\} = F_X(x_2) - F_X(x_1), \quad \text{for any } x_1 \leq x_2. \tag{3.11}$$

Property (iii) allows us to define the inverse distribution function, $F_X^{-1}(y)$, by

$$F_X^{-1}(y) = \min\{x: F_X(x) = y\}. \tag{3.12}$$

In words, since $F_X(x)$ may not be strictly increasing in x, $F_X^{-1}(y)$ is defined as the smallest value x, such that $F_X(x) = y$. The inverse distribution function is extensively used to generate realizations of random variables (see Chapter 4).

3.3.3 Probability Density Functions

If $F_X(x)$ is continuous and differentiable in x, then the associated *probability density function* (*pdf*), $f_X(x)$, is the derivative function

$$f_X(x) = \frac{d}{dx} F_X(x), \quad -\infty < x < \infty. \tag{3.13}$$

The pdf has the following properties for $-\infty < x < \infty$:

(i) $f_X(x) \geq 0$.

(ii) $F_X(x) = \int_{-\infty}^{x} f_X(x)\, dx$,

and in particular,

$$F_X(\infty) = \int_{-\infty}^{\infty} f_X(x)\, dx = 1.$$

Property (ii) implies the formula

$$\Pr\{x_1 \leq X \leq x_2\} = F_X(x_2) - F_X(x_1) = \int_{x_1}^{x_2} f_X(x)\, dx, \quad \text{for any } x_1 \leq x_2. \tag{3.14}$$

For a discrete random variable X, the associated pmf is sometimes referred to as a pdf as well. This identification is justified by the fact that a mathematical abstraction allows us, in fact, to define differencing as the discrete analog of differentiation. Indeed, for a discrete real-valued random variable X, we can write

$$F_X(x) = \sum_{y \leq x} f_X(y), \quad -\infty < x < \infty, \tag{3.15}$$

and each value, $f_X(x) = p_X(x)$, can be recovered by differencing in Eq. 3.15.

3.3.4 Joint Distributions

Let $X_1, X_2, \ldots, X_n$ be n real-valued random variables over a common probability space. The *joint cdf* of $X_1, X_2, \ldots, X_n$ is the function

$$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \Pr\{X_1 \leq x_1, \ldots, X_n \leq x_n\}, \; -\infty < x_i < \infty, \quad i = 1, \ldots, n. \tag{3.16}$$

Similarly, the *joint pdf*, when it exists, is obtained by multiple partial differentiation,

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \frac{\partial}{\partial x_1} \cdots \frac{\partial}{\partial x_n} F_{X_1,\ldots,X_n}(x_1, \ldots, x_n), \; -\infty < x_i < \infty, \quad i = 1, \ldots, n. \tag{3.17}$$

In this context, each cdf $F_{X_i}(x)$ and pdf $f_{X_i}(x)$ are commonly referred to as a *marginal distribution* and *marginal density*, respectively.

The random variables $X_1, X_2, \ldots, X_n$ are *mutually independent*, if

$$F_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = \prod_{i=1}^{n} F_{X_i}(x_i),\ -\infty < x_i < \infty,\quad i = 1,\ldots,n \tag{3.18}$$

or equivalently

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = \prod_{i=1}^{n} f_{X_i}(x_i),\ -\infty < x_i < \infty,\quad i = 1,\ldots,n, \tag{3.19}$$

provided that the densities exist. In other words, mutual independence is exhibited when joint distributions or densities factor out into their marginal components.

A set of random variables, $X_1, X_2, \ldots, X_n$, are said to be *iid* (*independently, identically distributed*), if they are mutually independent and each of them have the same marginal distribution.

3.4 EXPECTATIONS

The expectation of a random variable is a statistical operation that encapsulates the notion of "averaging." In other words, it assigns to a real-valued random variable, X, a number, $\mathrm{E}[X]$, called the *mean* or *expected value* or just the *expectation* of X. The expectation operation converts a random variable, X, to a deterministic scalar quantity, the mean value $\mathrm{E}[X]$, which can be thought of as a "central value" of X.

The mathematical definition of expectation varies according to the nature of the underlying state space. For a discrete random variable X with pmf $p_X(x)$, we define

$$\mathrm{E}[X] = \sum_{x \in S} x\, p_X(x), \tag{3.20}$$

and for a continuous random variable with pdf $f_X(x)$, we define

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} x\, f_X(x)\, dx. \tag{3.21}$$

We mention that the expectations in Eqs. 3.20 and 3.21 are only defined when the corresponding sum or integral exist. In either case, the averaging action of the expectation yields a weighted sum or integral, where the weights are probabilities or densities.

Let X and Y be random variables, whose expectations exist, and let a and b be real numbers. Then,

$$\mathrm{E}[a\,X + b\,Y] = a\,\mathrm{E}[X] + b\,\mathrm{E}[Y]. \tag{3.22}$$

Equation 3.22 shows that expectation is linear.

3.5 MOMENTS

Moments are expectations of the powers of a random variable. They provide information on the underlying distribution, and are sometimes used as parameters of particular distribution functions. Mathematically, the k-th moment of X is given by

$$m_k = \mathrm{E}[X^k], \quad k = 1, 2, \ldots \tag{3.23}$$

Thus, for $k = 1, m_1 = \mathrm{E}[X]$ is just the mean. The second moment, $m_2 = \mathrm{E}[X^2]$, is used to define the *variance* of X

$$\mathrm{V}[X] = \mathrm{E}[X^2] - \mathrm{E}^2[X], \tag{3.24}$$

which measures the *variability* or *dispersion* of the random variable on the real line in units of X^2. Unlike expectation, the variance operation is not linear, since

$$\mathrm{V}[aX + bY] = a^2\ \mathrm{V}[X] + b^2\ \mathrm{V}[Y] + 2ab\ \mathrm{Cov}[X, Y], \tag{3.25}$$

where

$$\mathrm{Cov}[X, Y] = \mathrm{E}[XY] - \mathrm{E}[X]\ \mathrm{E}[Y] \tag{3.26}$$

is the *covariance* of X and Y. The covariance of two random variables is a measure of association, indicating how two random variables "vary together." This topic will be covered in greater detail in Section 3.6, where a more useful measure of association will be presented as a normalized covariance. For now, we just point out that (3.26) readily shows the nonlinearity of the covariance, since

$$\mathrm{Cov}[aX, bY] = ab\ \mathrm{Cov}[X, Y]. \tag{3.27}$$

An alternative measure of variability or dispersion is the *standard deviation* of X,

$$\sigma[X] = \sqrt{\mathrm{V}[X]} = \sqrt{\mathrm{E}[X^2] - \mathrm{E}^2[X]}, \tag{3.28}$$

which is expressed in units of X.

The *squared coefficient of variation* of X, is the statistic

$$\mathrm{c}^2[X] = \frac{\mathrm{V}[X]}{\mathrm{E}^2[X]}, \tag{3.29}$$

which is yet another measure of the variability or dispersion of X, this time normalized to a unitless quantity.

While the number of moments is infinite (though not all may exist), only the first few moments are considered in practice. In particular, the third moment influences the *skewness* (departure from symmetry) of the distribution of X via the *coefficient of skewness*

$$v[X] = \frac{\mathrm{E}[(X - \mathrm{E}[X])^3]}{\sigma^3[X]}, \tag{3.30}$$

which is negative, zero, or positive, according as the distribution is left-skewed, symmetric, or right-skewed, respectively. In a similar vein, the fourth moment influences the *kurtosis* of the distribution of X,

$$\mathrm{k}[X] = \frac{\mathrm{E}[(X - \mathrm{E}[X])^4]}{\sigma^4[X]} - 3, \tag{3.31}$$

which measures the degree of "fatness" of the distribution tail relative to a normal distribution with the same standard deviation (it is negative, zero, or positive, according as the distribution is less "fat," equally "fat," or more "fat," respectively).

Note that knowledge of a finite number of moments does not determine a distribution, except in special cases. Furthermore, in pathological cases, moments can be infinite or mathematically undefined, depending on the shape of the underlying distribution.

3.6 CORRELATIONS

Let X and Y be two real-valued random variables over a common probability space. It is sometimes necessary to obtain information on the nature of the *association* (probabilistic relation) between X and Y, beyond dependence or independence.

A useful measure of statistical association between X and Y is their *correlation coefficient* (often abbreviated to just *correlation*), defined by

$$\rho(X, Y) = \frac{\mathrm{E}[XY] - \mathrm{E}[X]\,\mathrm{E}[Y]}{\sigma[X]\,\sigma[Y]}, \tag{3.32}$$

which is well defined whenever the corresponding standard deviations exist and are finite. Note that the numerator of Eq. 3.32 is precisely $\mathrm{Cov}[X, Y]$. The division by the standard deviations normalizes the covariance into a correlation coefficient, so that it is invariant under scaling, that is,

$$\rho(a\,X, bY) = \rho(X, Y), \tag{3.33}$$

unlike its covariance counterpart (see Eq. 3.27).

The correlation coefficient has the following properties:

1. $-1 \leq \rho(X, Y) \leq 1$.
2. If X and Y are *independent* random variables, then X and Y are *uncorrelated*, that is $\rho(X, Y) = 0$. However, the converse is false, namely, X and Y may be *uncorrelated* and *dependent*, simultaneously.
3. If Y is a (deterministic) linear function of X, that is, $Y = aX + b$, then
   If $a > 0$, then $\rho(X, Y) = 1$.
   If $a < 0$, then $\rho(X, Y) = -1$.

Property (3) above provides a clue into the operational meaning of the correlation coefficient as a measure of *linear dependence* between X and Y. More specifically, $\rho(X, Y)$ measures the linear covariation of X and Y as described below.

First, if $\rho(X, Y) > 0$, then X and Y are *positively correlated* random variables in the sense that their realizations tend to behave as follows:

1. When $X(\omega)$ is a relatively *large* realization, then $Y(\omega)$ tends to be a comparatively *large* realization simultaneously.
2. When $X(\omega)$ is a relatively *small* realization, then $Y(\omega)$ tends to be a comparatively *small* realization simultaneously.
3. When multiple pairs $(X(\omega), Y(\omega))$ are plotted as a graph, the points tend to arrange themselves in a band of a positive slope. The higher the correlation, the narrower is the band, until it becomes a line with a positive slope for $\rho(X, Y) = 1$.

Second, if $\rho(X, Y) < 0$, then X and Y are *negatively correlated* random variables in the sense that their realizations tend to behave as follows:

1. When $X(\omega)$ is a relatively *large* realization, then $Y(\omega)$ tends to be a comparatively *small* realization simultaneously.
2. When $X(\omega)$ is a relatively *small* realization, then $Y(\omega)$ tends to be a comparatively *large* realization simultaneously.
3. When multiple pairs $(X(\omega), Y(\omega))$ are plotted as a graph, the points tend to arrange themselves in a band of a negative slope. The higher the correlation, the narrower is the band, until it becomes a line with a negative slope for $\rho(X, Y) = -1$.

Third, if $\rho(X, Y) = 0$, then X and Y are *uncorrelated* random variables in the sense that there is no apparent linear relation between realization pairs $X(\omega)$ and $Y(\omega)$. When multiple pairs $(X(\omega), Y(\omega))$ are plotted on as a graph, the points form a "blob" with no apparent "direction."

Recall that correlation is a *weaker* concept than dependence, since it only measures linear dependence (X and Y may be related by another functional relation, e.g., quadratic). Still, linear dependence is a common instance of dependence, and is often taken as a *proxy* for dependence.

3.7 COMMON DISCRETE DISTRIBUTIONS

This section reviews the most commonly used discrete distributions and the underlying random experiment, and discusses their use in simulation modeling. For more information, see Bratley et al. (1987) or Law and Kelton (2000). We shall use indicator functions, defined for any set A by

$$1_A(x) = \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{if } x \notin A \end{cases}. \tag{3.34}$$

We also routinely indicate that a random variable X has distribution D by the notation $X \sim \mathrm{D}$.

3.7.1 Generic Discrete Distribution

A discrete random variable, X, corresponds to a trial (random experiment) with a countable (finite or infinite) number of distinct outcomes. Thus, its state space has the form $S = \{s_1, \ldots, s_i, \ldots\}$, where a generic state (realization), s_i, may be any symbol (it is common, however, to code the states by integers, when convenient). The generic *discrete distribution* is denoted by $\mathrm{Disc}(\{(p_i, v_i): i = 1, 2, \ldots\})$,[1] where each parameter pair, (p_i, v_i), corresponds to $\Pr\{X = v_i\} = p_i$.

The pmf of $X \sim \mathrm{Disc}(\{(p_i, v_i): i = 1, 2, \ldots\})$ is given by

[1] Note that while distribution names resemble those of Arena, the corresponding parameter definitions may differ from their Arena counterparts.

$$p_X(x) = \sum_{v_i \in S} 1_{\{v_i\}}(x)\, p_i = \begin{cases} p_i, & \text{if } x = v_i \text{ for some } i \\ 0, & \text{otherwise} \end{cases} \tag{3.35}$$

and for a real-valued state space, say $S = \{1, 2, \ldots\}$, the corresponding distribution function is given by

$$F_X(x) = \sum_{i=1}^{[x]} p_i = \begin{cases} 0, & \text{if } x < 1 \\ \sum_{i=1}^{k} p_i, & \text{if } k \le x < k+1 \end{cases} \tag{3.36}$$

where $[x]$ is the integral part of x.

The generic discrete distribution may be used to model a variety of situations, characterized by a discrete outcome. In fact, all other discrete distributions are simply useful specializations of the generic case.

3.7.2 Bernoulli Distribution

A *Bernoulli* random variable, X, corresponds to a trial with two possible outcomes: *success* or *failure*. Thus, its state space has the form $S = \{0, 1\}$, where state 0 codes for a *failure* realization and state 1 codes for a *success* realization. The Bernoulli distribution is denoted by Ber(p), where p represents the probability of success (and therefore, $1 - p$ is the probability of failure).

The pmf of $X \sim \text{Ber}(p)$ is

$$p_X(k) = \begin{cases} p, & k = 1 \\ 1 - p, & k = 0 \end{cases} \tag{3.37}$$

and the corresponding mean and variance are given by the formulas:

$$\text{E}[X] = p \tag{3.38}$$

and

$$\text{V}[X] = p(1 - p). \tag{3.39}$$

A Bernoulli random variable may be used to model whether a job departing from a machine is defective (failure) or not (success).

3.7.3 Binomial Distribution

A *binomial* random variable, $X = \sum_{k=1}^{n} X_k$, is the sum of n independent Bernoulli random variables, X_k, with a common success probability, p. Thus, its state space has the form $S = \{0, 1, \ldots, n\}$, and state k corresponds to a realization of k successes in n Bernoulli trials. The binomial distribution is denoted by B(n, p).

The pmf of $X \sim \text{B}(n, p)$ is

$$p_X(k) = \binom{n}{k} p^k (1 - p)^{n-k}, k = 0, 1, \ldots, n, \tag{3.40}$$

where $\binom{n}{k} = \frac{n!}{k!(n-k)!}$, $n \geq k \geq 0$, and the corresponding mean and variance are given by the formulas

$$\mathrm{E}[X] = np \tag{3.41}$$

and

$$\mathrm{V}[X] = np(1-p). \tag{3.42}$$

A binomial random variable may be used to model the total number of defective items in a given batch. Such a binomial trial can be a much faster procedure than conducting multiple Bernoulli trials (for each item separately).

3.7.4 Geometric Distribution

A *geometric* random variable, X, is the number of Bernoulli trials to and including the first success. The geometric distribution is denoted by Ge(p), where p represents the probability of success (and therefore, $1-p$ is the probability of failure). Since the number of trials is potentially unbounded, the state space becomes $S = \{1, 2, \ldots, k, \ldots\}$.

The pmf of $X \sim \mathrm{Ge}(p)$ is

$$p_X(k) = (1-p)^{k-1}p, k = 1, 2, \ldots \tag{3.43}$$

and the corresponding mean and variance are given by the formulas

$$\mathrm{E}[X] = \frac{1}{p} \tag{3.44}$$

and

$$\mathrm{V}[X] = \frac{1-p}{p^2}. \tag{3.45}$$

A geometric random variable may be used to model the number of good product units, separating consecutive bad (defective) ones.

The geometric distribution is also widely used in mathematical models, because it often renders the analysis tractable. This tractability is due to the fact that the geometric distribution is the only discrete distribution with the so-called *memoryless property*, namely,

$$\Pr\{X > k+n | X > k\} = \Pr\{X > n\}, \quad \text{for all } k, n \geq 1. \tag{3.46}$$

This equation states that the probability that the remaining number of trials to the next success is independent of the number of trials elapsed since the previous success.

3.7.5 Poisson Distribution

A *Poisson* random variable, X, can be thought of as a generalization of a binomial random variable from discrete trials to continuous trials. It represents the total number

of successes as the limit of a sequence of binomial trials, in which n tends to infinity and p tends to 0, such that the product $np = \lambda$ is fixed and represents the rate of successes per time unit. The resulting Poisson random variable then represents the number of successes in a unit interval. Since the number of successes is potentially unbounded, the state space becomes $S = \{0, 1, \ldots, k, \ldots\}$. The Poisson distribution is denoted by Pois(λ).

The pmf of $X \sim \text{Pois}(\lambda)$ is

$$p_X(k) = \begin{cases} \frac{e^{-\lambda}\lambda^k}{k!}, & k = 0, 1, \ldots \\ 0, & \text{otherwise} \end{cases} \tag{3.47}$$

and the corresponding mean and variance are given by

$$\text{E}[X] = \lambda \tag{3.48}$$

and

$$\text{V}[X] = \lambda \tag{3.49}$$

A Poisson random variable is often used to model the number of random occurrences in a time interval. Examples include the number of machine failures in a time interval, number of customer demands in a time interval, and so on.

3.8 COMMON CONTINUOUS DISTRIBUTIONS

This section reviews the most commonly used continuous distributions and the underlying random experiment, and discusses their use in simulation modeling. For more information, see Bratley et al. (1987) or Law and Kelton (2000).

3.8.1 Uniform Distribution

A *uniform* random variable, X, assumes values in an interval $S = [a, b], b > a$, such that each value is equally likely. The uniform distribution is denoted by Unif(a, b), and is the simplest continuous distribution.

The pdf of $X \sim \text{Unif}(a, b)$ is

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & \text{if } a \le x \le b\ , \\ 0, & \textit{otherwise} \end{cases} \tag{3.50}$$

and the cdf is

$$F_X(x) = \begin{cases} 0, & \text{if } x < a \\ \frac{x-a}{b-a}, & \text{if } a \le x \le b \\ 1, & \text{if } x > b. \end{cases} \tag{3.51}$$

The corresponding mean and variance are given by the formulas

$$\text{E}[X] = \frac{a+b}{2} \tag{3.52}$$

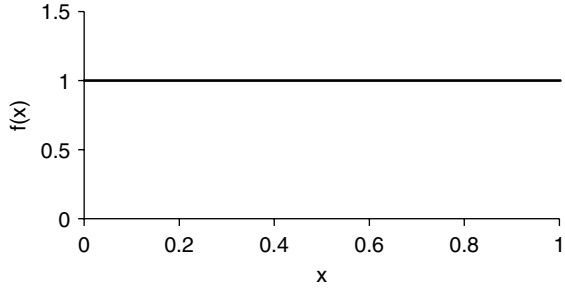

Figure 3.1 Density function of the Unif(0, 1) distribution.

and

$$\mathrm{V}[X] = \frac{(b-a)^2}{12}. \tag{3.53}$$

A graph of the pdf of a uniform distribution is depicted in Figure 3.1.

A uniform random variable is commonly employed in the absence of information on the underlying distribution being modeled.

3.8.2 Step Distribution

A *step* or *histogram* random variable, X, generalizes the uniform distribution in that it constitutes a *probabilistic mixture* of uniform random variables. The step distribution is denoted by $\mathrm{Cont}(\{(p_j, l_j, r_j): j = 1, 2, \ldots, J\})$, where the parameters have the following interpretation: $X \sim \mathrm{Unif}(l_j, r_j)$ with probability p_j, $j = 1, 2, \ldots, J$. Thus, the state space of X is the union of intervals,

$$S = \bigcup_{j=1}^{J} [l_j, r_j).$$

The pdf of $X \sim \mathrm{Cont}(\{(p_j, l_j, r_j): j = 1, 2, \ldots, J\})$ is given by

$$f_X(x) = \sum_{j=1}^{J} 1_{[l_j, r_j)}(x) \frac{p_j}{r_j - l_j} = \begin{cases} \frac{p_j}{r_j - l_j}, & \text{if } l_j \le x < r_j \\ 0, & \textit{otherwise} \end{cases} \tag{3.54}$$

Thus, the resulting pdf is a step function (mixture of uniform densities) as illustrated in by Figure 3.2, and the corresponding cdf is given by

$$F_X(x) = \begin{cases} 0, & \text{if } x < l_1 \\ \sum_{j=1}^{J} 1_{[l_j, r_j)}(x) \left[\sum_{i=1}^{j-1} p_i + (x - l_j) \frac{p_j}{r_j - l_j} \right], & \text{if } l_1 \le x < r_J \\ 1, & \text{if } x \ge r_J. \end{cases} \tag{3.55}$$

The corresponding mean and variance are given by the formulas

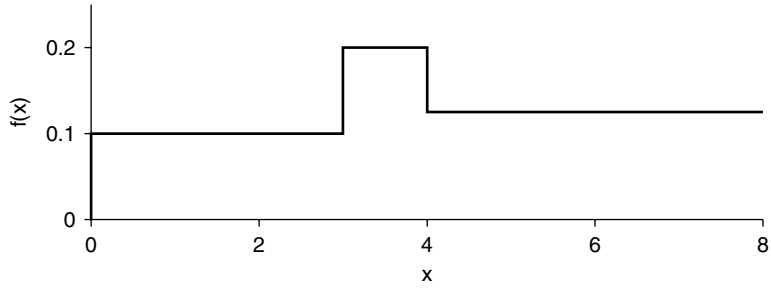

Figure 3.2 Density function of the Cont({(0.3, 0, 3), (0.2, 3, 4), (0.5, 4, 8)}) distribution.

$$\mathrm{E}[X] = \sum_{j=1}^{J} p_j \frac{l_j + r_j}{2} \tag{3.56}$$

and

$$\mathrm{V}[X] = \frac{1}{3}\sum_{j=1}^{J} p_j \left(l_j^2 + l_j r_j + r_j^2\right) - \frac{1}{4}\left(\sum_{j=1}^{J} p_j (r_j + l_j)\right)^2 \tag{3.57}$$

A step random variable is routinely used to model an empirical distribution, estimated by a histogram. Suppose the histogram has J cells. Then cell j coincides with the interval $[l_j, r_j)$, and the probability estimate (relative frequency) of the cell will be assigned as the value of the corresponding p_j.

3.8.3 Triangular Distribution

A *triangular* random variable, X, assumes values in an interval $S = [a, b]$, with the most "likely" value (the *mode*) being some point $c \in [a, b]$. The likelihood increases linearly in the subinterval $[a, c]$, and decreases linearly in the subinterval $[c, b]$, so that the density has a triangular shape (see Figure 3.3). The triangular distribution is denoted by Tria(a, c, b).

The pdf of $X \sim \mathrm{Tria}(a, c, b)$ is

$$f_X(x) = \begin{cases} \frac{2(x-a)}{(b-a)(c-a)}, & \text{if } a \le x \le c \\ \frac{2(b-x)}{(b-a)(b-c)}, & \text{if } c \le x \le b \\ 0, & \textit{otherwise.} \end{cases} \tag{3.58}$$

The corresponding mean and variance are given by the formulas

$$\mathrm{E}[X] = \frac{a + b + c}{3} \tag{3.59}$$

and

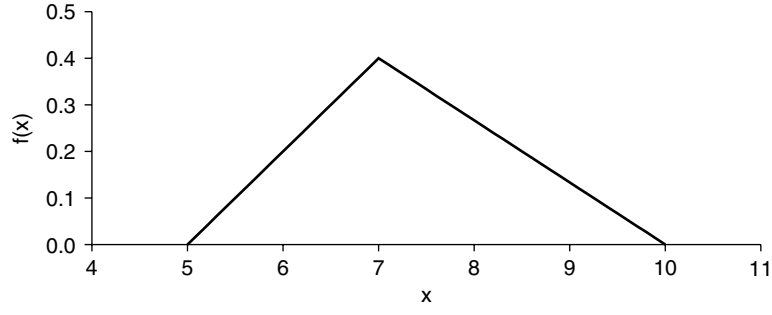

Figure 3.3 Density function of the Tria(5, 7, 10) distribution.

$$\mathrm{V}[X] = \frac{a^2 + b^2 + c^2 - ab - ac - bc}{18} \tag{3.60}$$

A triangular random variable is used when the underlying distribution is unknown, but it is reasonable to assume that the state space ranges from some minimal value, a, to some maximal value, b, with the most likely value being somewhere in between, at c. The choice of c then determines the skewness of the triangular distribution. The piecewise linear form of the pdf curve of Figure 3.3 is the simplest way to represent this kind of behavior.

3.8.4 Exponential Distribution

An *exponential* random variable, X, assumes values in the positive half-line $S = [0, \infty]$. The exponential distribution is denoted by Expo(λ), where λ is called the *rate* parameter.[2]

The pdf of $X \sim \mathrm{Expo}(\lambda)$ is

$$f_X(x) = \lambda e^{-\lambda x}, \quad x \geq 0, \tag{3.61}$$

and the cdf is

$$F_X(x) = 1 - e^{-\lambda x}, \quad x \geq 0. \tag{3.62}$$

The corresponding mean and variance are given by the formulas

$$\mathrm{E}[X] = \frac{1}{\lambda} \tag{3.63}$$

and

$$\mathrm{V}[X] = \frac{1}{\lambda^2}. \tag{3.64}$$

A graph of the pdf of an exponential distribution is depicted in Figure 3.4.

[2] Note that in Arena, the corresponding parameter is the mean $1/\lambda$, rather than the rate λ.

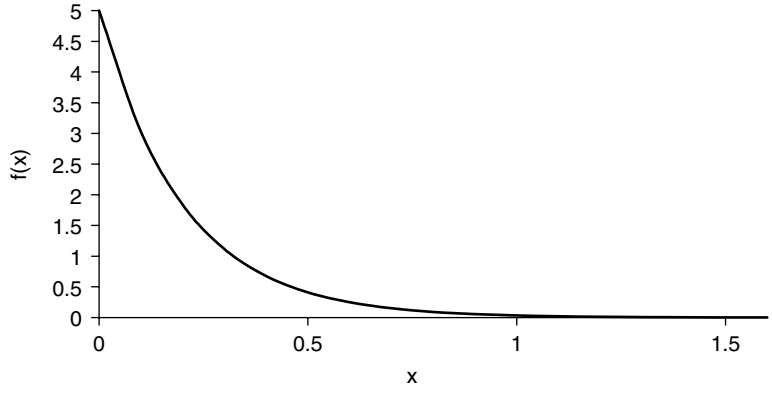

Figure 3.4 Density function of the Expo(5) distribution.

Exponential random variables are widely used to model "random" interarrival times in continuous time, especially when these are iid. Examples include customer interarrivals, times to failure, and so on.

The exponential distribution is also widely used in mathematical models, because it often renders the analysis tractable. This tractability is due to the fact that the exponential distribution is the only continuous distribution with the so-called *memoryless property*, namely,

$$\Pr\{X > s + t | X > s\} = \Pr\{X > t\}, \text{for all } s, t \geq 0. \tag{3.65}$$

The equation above states that the probability that the remaining time to the next arrival is independent of the time elapsed since the previous arrival. In fact, the exponential distribution constitutes a generalization of the geometric distribution to continuous time.

3.8.5 Normal Distribution

In Arena, mean and standard deviation are used as input parameters

A *normal* random variable, X, can assume any value on the real line $S = (-\infty, \infty)$. The normal distribution is denoted by Norm(μ, σ^2), where μ is the mean (*scale* parameter) and σ^2 is the variance (*shape* parameter), and has the familiar bell shape (Figure 3.5), popularly known as the *bell curve*. In the technical literature, it is also known as the *gaussian* distribution, as a tribute to the mathematician Gauss. The special case Norm(0,1) is known as the *standard normal* distribution. Another transformation of normal random variables, implemented by Arena, results in the so-called *Johnson* distribution (see Kelton et. al. 1998).

The pdf of $X \sim \text{Norm}(\mu, \sigma^2)$ is

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x \leq \infty. \tag{3.66}$$

The corresponding mean and variance are given by the formulas

$$\text{E}[X] = \mu \tag{3.67}$$

and

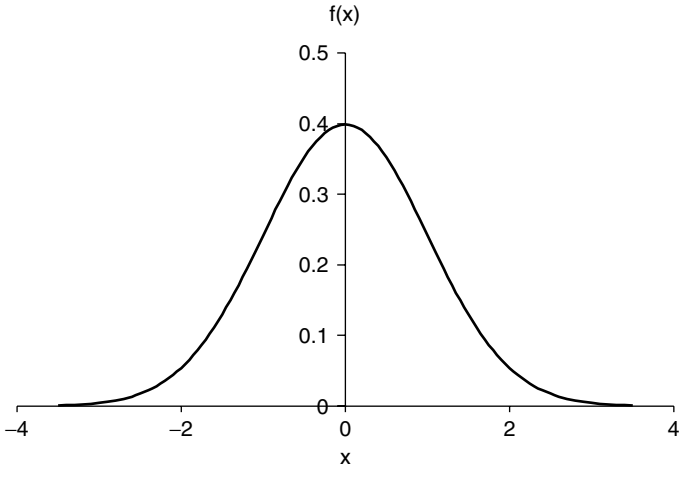

Figure 3.5 Density function of the Norm(0, 1) distribution.

$$\mathrm{V}[X] = \sigma^2. \tag{3.68}$$

A graph of the pdf of the standard normal distribution is depicted in Figure 3.5.

An important property of normal random variables is that they can always be standardized. This means that if $X \sim \mathrm{Norm}(\mu, \sigma^2)$, then

$$Z = \frac{X - \mu}{\sigma^2} \sim \mathrm{Norm}(0, 1)$$

is a standard normal random variable. Furthermore, if $X \sim \mathrm{Norm}(\mu_X, \sigma_X^2)$ and $Y \sim \mathrm{Norm}(\mu_Y, \sigma_Y^2)$ are independent normal variables, then

$$aX + bY \sim \mathrm{Norm}(a\,\mu_X + b\,\mu_Y, a^2\sigma_X^2 + b^2\sigma_Y^2),$$

which shows the linearity of normal distributions.

A normal random variable is used to model many random phenomena that can be expressed as sums of random variables, by virtue of the *central limit theorem*. This fundamental theorem asserts that the distribution of the sum approaches the normal distribution when the addends are iid (and in other cases as well).

The analyst should be careful in using normal distributions to model random phenomena, which cannot assume negative values (e.g., interarrival times). If the mean, μ, is large enough, then a negative value would be sampled relatively rarely, and may be simply ignored until further sampling yields a "legal" non-negative value. The analyst should be aware, however, that this procedure samples from a distribution that is no longer normal; rather, it is a normal distribution, conditioned on the outcome being non-negative.

3.8.6 Lognormal Distribution

A *lognormal* random variable, X, assumes values in the positive half-line $S = [0, \infty]$. The lognormal distribution is denoted by $\mathrm{Logn}(\mu, \sigma)$, where μ is a *scale* parameter and σ is a *shape* parameter.

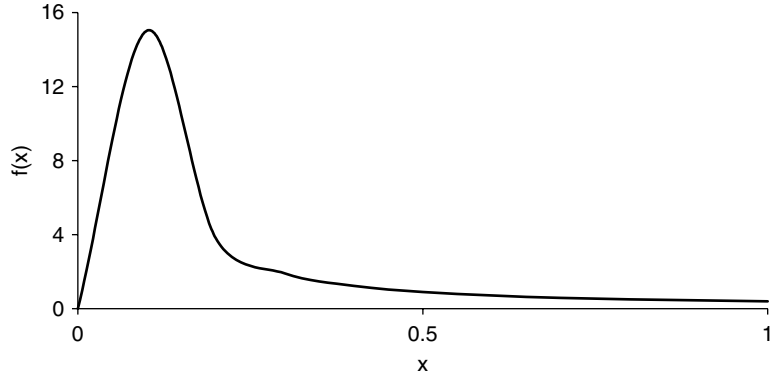

Figure 3.6 Density function of the Logn(0, 1) distribution.

The pdf of $X \sim \text{Logn}(\mu, \sigma)$ is

$$f_X(x) = \frac{1}{\sigma x\sqrt{2\pi}} e^{\frac{(\ln x - \mu)^2}{2\sigma^2}}, \quad x \geq 0. \tag{3.69}$$

The corresponding mean and variance are given by the formulas

$$\text{E}[X] = e^{\mu + \sigma^2/2} \tag{3.70}$$

and

$$\text{V}[X] = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1). \tag{3.71}$$

A graph of the pdf of a lognormal distribution is shown in Figure 3.6.

A lognormal random variable, X, can be represented as $X = e^Y$, where $Y \sim \text{Norm}(\mu, \sigma^2)$. It is always positive, and is often used in finance to model financial random processes.

3.8.7 Gamma Distribution

A *gamma* random variable, X, assumes values in the positive half-line $S = [0, \infty]$. The gamma distribution is denoted by $Gamm(\alpha, \beta)$, where $\alpha > 0$ is the *shape* parameter and $\beta > 0$ is the *scale* parameter.[3]

The pdf of $X \sim \text{Gamm}(\alpha, \beta)$ is

$$f_X(x) = \frac{x^{\alpha-1}e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)}, \quad x \geq 0, \tag{3.72}$$

where

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1}e^{-y}\,dy \tag{3.73}$$

[3] Note that the gamma distribution in Arena has the parameters in reverse order. That is to say, Arena uses notation $\text{Gamm}(\beta, \alpha)$, while the pdf is the same as (3.72).

is known as the *gamma function*. The corresponding mean and variance are given by the formulas

$$\mathrm{E}[X] = \alpha\beta \tag{3.74}$$

and

$$\mathrm{V}[X] = \alpha\beta^2. \tag{3.75}$$

Three graphs of the pdf of gamma distributions are depicted in Figure 3.7.

As the parameter names suggest, the gamma distribution is a parameterized family of distributions. A particular distribution can be selected with an appropriate choice of the shape and scale parameters. For example, for $\alpha = 1$ and $\beta = 1/\lambda$, we obtain the exponential distribution Expo(λ), since $\Gamma(1) = 1$. More generally, for integer $\alpha = k \geq 1$ and $\beta = 1/\lambda$, we obtain an *Erlang* distribution, denoted by Erl(k, λ),[4] and given by

$$f_X(x) = \frac{\lambda^k x^{k-1} e^{-\lambda x}}{(k-1)!}, \quad x \geq 0. \tag{3.76}$$

The Erlang distribution is useful because an Erlang random variable can be represented as the sum of k iid exponential random variables, with a common rate, λ, and in particular, Erl($1, \lambda$) = Expo(λ). An Erlang random variable is useful in modeling multiple exponential "phases" with a common rate. For example, the model of a manufacturing subsystem, where products are serially processed without waiting in k processes with common processing rate λ, can be equivalently aggregated into one process with service distribution Erl(k, λ).

Another useful specialization is obtained for $\alpha = n/2$ (n even) and $\beta = 2$, which is called the *chi-square* distribution with n degrees of freedom, and denoted by $\chi^2(n)$. A $\chi^2(n)$ distributed random variable, X, can be represented as a sum

$$X = \sum_{i=1}^{n} Y_i^2$$

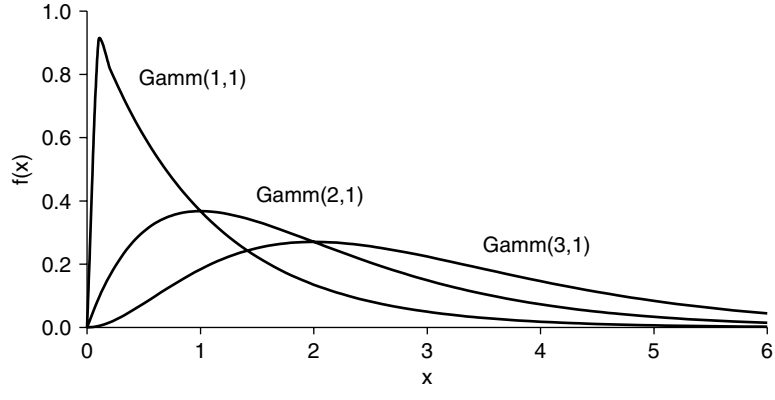

Figure 3.7 Density functions of the Gamm(1, 1), Gamm(2, 1), and Gamm(3, 1) distributions.

[4] Note that the Erlang distribution in Arena is represented with slightly different parameters. Arena uses Erl($1/\lambda, k$).

of n independent squared standard normal random variables, Y_i. The class of *chi-square* distributed random variables has extensive applications in statistics.

3.8.8 Student's t Distribution

A *Student's t* random variable, X, (t random variable, for short) can assume any value on the real line $S = (-\infty, \infty)$. The t distribution is denoted by t(n), where the n parameter is the *number of degrees of freedom*.

The pdf of $X \sim \text{t}(n),\ n > 2$, is

$$f_X(x) = \frac{\Gamma((n+1)/2)}{\sqrt{\pi n}\ \Gamma(n/2)} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2}, \quad -\infty \leq x \leq \infty \tag{3.77}$$

where Γ is the gamma function of Eq. 3.73. The corresponding mean and variance are given by the formulas

$$\text{E}[X] = 0 \tag{3.79}$$

and

$$\text{V}[X] = \frac{n}{n-2}. \tag{3.80}$$

A graph of the pdf of a Student's t distribution is depicted in Figure 3.8.

A t(n) distributed random variable X can be represented as

$$X = \frac{Z}{\sqrt{Y/n}}, \tag{3.81}$$

where $Z \sim \text{Norm}(0, 1)$ is a standard normal random variable, $Y \sim \chi^2(n)$ is a chi-square random variable with n degrees of freedom, and Z and Y are independent. As can be seen in Figure 3.8, t(n) distributions have a functional form similar to that of the standard normal distribution, Norm(0, 1), but with "fatter" tails, which give rise to larger variances as indicated by Eq. 3.80. However, as the degrees-of-freedom parameter, n, tends to infinity, the t(n) distribution converges to Norm(0, 1).

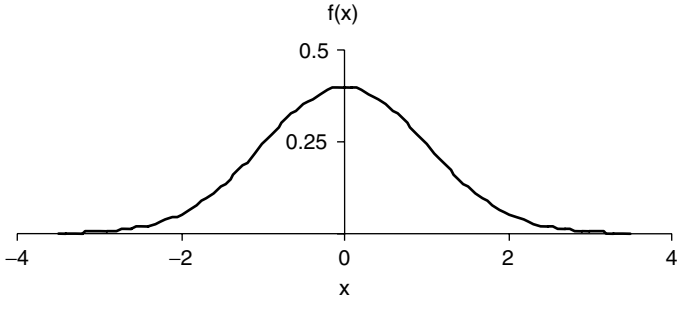

Figure 3.8 Density function of the t(10) distribution.

3.8.9 F Distribution

An F random variable, X, assumes values in the positive half-line $S = [0, \infty]$. The F distribution is denoted by $\mathrm{F}(n_1, n_2)$, where n_1 and n_2 are the *degrees of freedom* parameters.

The pdf of $X \sim \mathrm{F}(n_1, n_2)$ is

$$f_X(x) = \frac{\Gamma((n_1+n_2)/2)}{\Gamma(n_1/2)\Gamma(n_2/2)} \left[\frac{n_1}{n_2}\right]^{n_1/2} \frac{x^{(n_1/2)-1}}{\left[1+\frac{n_1}{n_2}x\right]^{(n_1+n_2)/2}}, \; 0 \le x \le \infty \tag{3.82}$$

where Γ is the gamma function of Eq. 3.73. The corresponding mean and variance are given by the formulas

$$\mathrm{E}[\boldsymbol{X}] = \frac{n_2}{n_2 - 2} \quad (\text{for } n_2 > 2) \tag{3.83}$$

and

$$\mathrm{V}[\boldsymbol{X}] = \frac{2n_2^2(n_1+n_2-2)}{n_1(n_2-4)(n_2-2)} \quad (\text{for } n_2 > 4). \tag{3.84}$$

An $\mathrm{F}(n_1, n_2)$ density is depicted in Figure 3.9.

An $\mathrm{F}(n_1, n_2)$ distributed random variable X can be represented as

$$X = \frac{V/n_1}{W/n_2}, \tag{3.85}$$

where $V \sim \chi^2(n_1)$ and $W \sim \chi^2(n_2)$ are independent chi-square random variables with the corresponding degrees of freedom. The $\mathrm{F}(n_1, n_2)$ distribution is skewed to the right, but it becomes less skewed as the degrees-of-freedom parameters, n_1 and n_2, increase in magnitude.

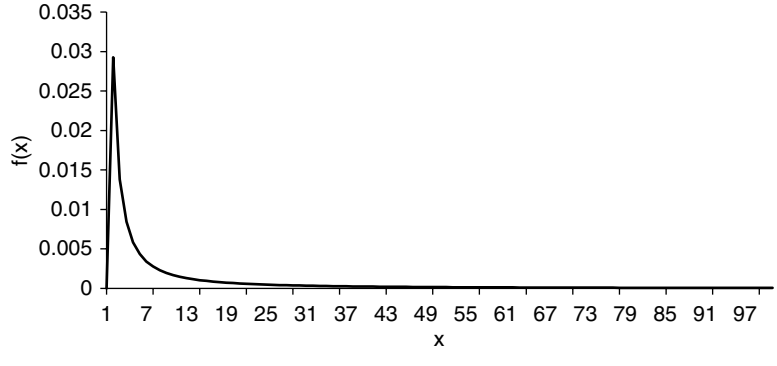

Figure 3.9 Density function of the F(1, 1) distribution.

3.8.10 Beta Distribution

A *beta* random variable, X, assumes values in the unit interval $S = [0, 1]$, although it may be scaled and shifted to any interval. The beta distribution is denoted by Beta(α, β), where $\alpha > 0$ and $\beta > 0$ are two *shape* parameters.[5]

The pdf of $X \sim \text{Beta}(\alpha, \beta)$ is

$$f_X(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}, \quad 0 \le x \le 1, \tag{3.86}$$

where

$$B(\alpha, \beta) = \int_0^\infty y^{\alpha-1}(1-y)^{\beta-1}dy = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \tag{3.87}$$

is known as the *beta* function, and is defined in terms of the gamma function of Eq. 3.73. The corresponding mean and variance are given, respectively, by the formulas

$$\text{E}[X] = \frac{\alpha}{\alpha+\beta} \tag{3.88}$$

and

$$\text{V}[X] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}. \tag{3.89}$$

Three graphs of the pdf of beta distributions are depicted in Figure 3.10.

A beta random variable is often used in statistics to model an unknown probability, regarded as a random variable.

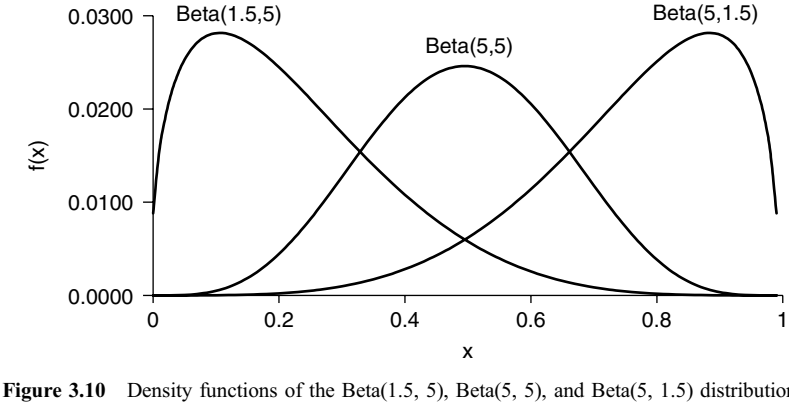

Figure 3.10 Density functions of the Beta(1.5, 5), Beta(5, 5), and Beta(5, 1.5) distributions.

[5] ~~Note that the beta distribution in Arena has the parameters in reverse order.~~
Arena uses BETA(Beta, Alpha), but the Beta is the α in (3.86), and the Alpha is the β in (3.86).

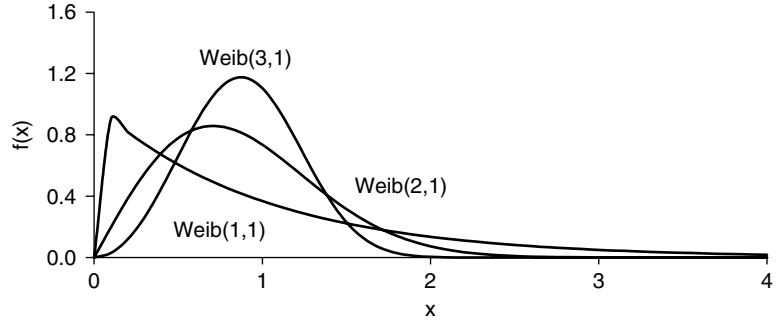

Figure 3.11 Density functions of the Weib(1, 1), Weib(2, 1), and Weib(3, 1) distributions.

3.8.11 Weibull Distribution

A *Weibull* random variable, X, assumes values in the positive half-line $S = [0, \infty]$. The Weibull distribution is denoted by Weib(α, β), where $\alpha > 0$ is the *shape* parameter and $\beta > 0$ is the *scale* parameter.[6]

The pdf of $X \sim$ Weib(α, β) is

$$f_X(x) = \frac{\alpha}{\beta^\alpha} x^{\alpha-1} e^{-(x/\beta)^\alpha}, \quad x \geq 0. \tag{3.90}$$

The corresponding mean and variance, respectively, are given by the formulas

$$\mathrm{E}[X] = \alpha\Gamma(1/\beta + 1) \quad \text{(struck)}$$
$$\mathrm{E}[X] = \beta\Gamma(1/\alpha + 1) \tag{3.91}$$

and

$$\mathrm{V}[X] = \alpha^2[\Gamma(2/\beta + 1) - \Gamma^2(1/\beta + 1)] \quad \text{(struck)}$$
$$\mathrm{V}[X] = \beta^2[\Gamma(2/\alpha + 1) - \Gamma^2(1/\alpha + 1)] \tag{3.92}$$

in terms of the gamma function of Eq. 3.73. Three graphs of the pdf of Weibull distributions are depicted in Figure 3.11.

The Weibull distribution is a parametric family of distributions. For $\alpha = 1$ and $\beta = 1/\lambda$, it becomes the exponential Expo(λ) distribution, while for $\alpha = 2$, it becomes the *Rayleigh* distribution (often used in artillery trajectory computations). Weibull random variables are often used in modeling the aging process of components in reliability analysis.

3.9 STOCHASTIC PROCESSES

A *stochastic process* is a time-indexed set of random variables, $\{X_t\}_{t \in T}$, with a common state space S, over a common probability space. The associated probability measure is called the *probability law* of the process. The time set, T, can be discrete or continuous, typically of the form $T = \{0, 1, \ldots, n, \ldots\}$ or $T = [0, \Theta]$, where Θ is either finite or infinite. For example, X_t may model the inventory level of a particular product in a warehouse at time t.

[6] Note that the Weibull distribution in Arena has the parameters in reverse order. That is to say, Arena uses notation Weib(β, α), while the pdf is the same as (3.90).

Stochastic processes are widely used to model random phenomena that evolve in time, such as arrival streams, service times, and routing decisions, to name but a few. In fact, simulation runs (replications) typically generate extensive realizations of multiple interacting stochastic processes. A realization of a stochastic process is also called a *sample path*, because it represents a possible history of its constituent time-indexed random variables. Most processes generated by simulation, but by no means all, are *stationary processes*, that is, their joint distributions (of any dimension) do *not* change in time.

The *autocorrelation function* of a stochastic process is the correlation coefficient of its lagged random variables,

$$\rho(\tau, \delta) = \frac{\mathrm{E}[X_\tau X_{\tau+\delta}] - \mathrm{E}[X_\tau]\mathrm{E}[X_{\tau+\delta}]}{\sigma[X_\tau]\sigma[X_{\tau+\delta}]}, \quad \tau \in T, \delta \geq 0. \tag{3.93}$$

For stationary processes, the autocorrelation function depends only on the first argument, τ. The autocorrelation function is often used as a convenient proxy for temporal dependence in stochastic processes.

The next few subsections discuss several stochastic processes, commonly used in simulation. Generation of their sample paths is discussed in Chapter 4. For further reading on stochastic processes, we recommend that the reader to refer to Ross (1993), and Taylor and Karlin (1984).

3.9.1 Iid Processes

Independent identically distributed (*iid*) processes have the simplest possible probability law, since all random variables indexed by its time set are mutually independent and share a common marginal distribution. This means that iid processes do not have *temporal dependence* (*time dependence*) in the sense that their "past" is always probabilistically irrelevant to their "future."

Iid processes are extensively used in simulation modeling, when justified by modeling considerations, or as a simplifying assumption in the absence of additional information. Typical examples are arrival processes, whose interarrival times are modeled as iid random variables or times to failure in a machine, which are often assumed to be iid.

3.9.2 Poisson Processes

A *Poisson process* $\{K_t\}_{t\geq 0}$ is a *counting process*, that is, it has state space $S = \{0, 1, \ldots\}$, continuous time set T, and nondecreasing sample paths; however, count increments may not exceed 1 (multiple simultaneous arrivals are not allowed). A random variable K_t from a Poisson process represents the (cumulative) *count* of some abstract "arrivals"; the last term actually connotes any phenomenon that can be declared to take place at discrete time points (e.g., job arrivals, failures, etc.). The distinguishing feature of any Poisson process is the *independent increment property*, which in its simplest form states that

$$\Pr\{K_{t+u} - K_t | K_s, s \leq t\} = \Pr\{K_{t+u} - K_t\}, \quad \text{for all } t, u \geq 0. \tag{3.94}$$

In words, a count increment in a future interval is independent of any past counts. It can be shown that this property alone forces the Poisson process to have a specific count increment distribution, and a specific interarrival distribution as follows:

1. Any count increment of the form $K_{t+u} - K_t$, over the interval $[t, t+u]$ of length u, has the Poisson distribution Pois(λu), for some $\lambda > 0$.
2. The interarrival times between successive arrivals are iid exponential with the aforementioned parameter, λ, that is, their distribution is Expo(λ).

In fact, conditions 1 and 2 are equivalent characterizations of the Poisson process. The parameter λ is the arrival rate of the Poisson process (expected number of arrivals per time unit).

The following operations on Poisson processes result in new Poisson processes (*closure* properties):

1. The *superposition* of independent Poisson processes (merging all their arrival points along the timeline) results in a new Poisson process. More specifically, if $\{K_t\}_{t\geq 0}$ and $\{L_t\}_{t\geq 0}$ are independent Poisson processes, with respective arrival rates λ_K and λ_L, then the superposition process, $\{K_t + L_t\}_{t\geq 0}$ is a Poisson process of rate $\lambda_K + \lambda_L$.
2. The *thinning* of a Poisson process (random deletion of its arrival points) results in a new Poisson process. More specifically, if $\{K_t\}_{t\geq 0}$ is a Poisson process of rate λ_K, from which arrivals are deleted according to independent Bernoulli trials with probability $1 - p$, then the thinned process, $\{L_t\}_{t\geq 0}$ is a Poisson process of rate $\lambda_L = p\lambda_K$.

The simplicity of Poisson processes and their closure properties render them a popular traffic model in network systems, because traffic merging and thinning of Poisson processes (by splitting such a stream into substreams) result in new Poisson processes. Moreover, Poisson processes have been widely used to model external arrivals to a variety of systems, where arriving customers make "independent arrival decisions." For example, telephone customers do not normally "coordinate" their phone calls, and customer demand arrivals are usually independent of each other. In these cases, the Poisson process assumption on the respective arrival processes may well be justified.

3.9.3 Regenerative (Renewal) Processes

A stochastic process $\{X_\tau : \tau \geq 0\}$ (discrete time or continuous time) is characterized as *regenerative* or *renewal* if it has (usually random) time points $T_1, T_2, \ldots$, such that the partial process histories $\{X_\tau : T_j \leq \tau < T_{j+1}\}$ over the intervals $[T_j, T_{j+1})$ are iid. In other words, the partial histories are independent statistical replicas of each other. For this reason, the time points $T_1, T_2, \ldots$ are referred to as *regeneration points* or *renewal points*, because they identify when the underlying process "renews" or "regenerates" itself statistically.

3.9.4 Markov Processes

Markov processes form the simplest class of dependent stochastic processes, with dependence extending only to the most "recent" past information. Formally,

$\{M_t\}_{t\geq 0}$ is a Markov process, if for all events $\{M_u \in A\}$, it satisfies the Markovian condition

$$\Pr\{M_u \in A | M_s : s \leq t\} = \Pr\{M_u \in A | M_t\} \text{ for all } 0 \leq t \leq u. \tag{3.95}$$

The probability law 3.95 is relatively simple. It stipulates that the probability of a future event probability $\Pr\{M_u \in A\}$ conditioned on past random variables $\{M_s : s \leq t\}$ (history) before u, equals the probability of the same future event, conditioned only on the most recent random variable, M_t. This means that in Markov processes, knowledge of information strictly predating some time t is immaterial for event probabilities after t.

The relative simplicity of Markov processes renders them popular models in analysis as well, without sacrificing the feature of temporal dependence. For example, discrete-time Markov processes with a discrete space S, known as *Markov chains*, are particularly simple. For a Markov chain, Eq. 3.95 becomes a matrix $Q = [q_{i,j}]$, called the *transition probability matrix*, where

$$q_{i,j} = \Pr\{M_{k+1} = j | M_k = i\} \text{ for any pair of states } i,j \in S. \tag{3.96}$$

The statistics of Markov chains can then be computed using matrix calculus.

Discrete-state Markov processes in continuous time are often classified as *jump processes*, because their sample paths have the form of step functions, whose discontinuities (jumps) correspond to state transitions. Markov jump processes have a simple structure that facilitates their generation:

1. Jumps are governed by transition probabilities similar to Eq. 3.96. The sequence of states visited by jumps is called the *jump chain*.
2. The time elapsed in state i is distributed exponentially with parameter λ_i, which depends only on state i but *not* on the state transitioned to.

As an example, consider an M/M/1 queue (iid exponential interarrival and service times), with the Markovian state being the number of customers in the system. The state jumps up and down following customer arrivals and service completions, respectively, and is otherwise constant.

A useful generalization of this kind of Markov jump processes is the class of *Markov renewal processes*. Here, the step-function form of sample paths is retained, as well as the Markovian structure of the jump chain. However, the times separating jumps can have a general (not necessarily exponential) distribution, and the transition probabilities of the time intervals separating successive jumps depend not only on the state jumped from, but also on the state jumped to. For more details, see Çinlar (1975).

3.10 ESTIMATION

An *estimator* is a random statistic, namely, a function of some observed random sample of data. A value of the estimator is called an *estimate* (usually the estimated quantity is some unknown parameter value). Note that an estimator is a random variable, while an estimate is one of its realizations. Good estimators are *unbiased*, that is, as the sample size grows to infinity, the expectations of such estimators (which are random variables) converge to the true parameter value, whatever it is.

Moments and related statistics are routinely estimated from sample data, using statistical estimators. Consider a (finite) sample $\vec{Y} = \{Y_1, Y_2, \ldots, Y_N\}$, where the random observations $\{Y_j,\ j = 1, \ldots, N\}$ have a common distribution with mean μ and variance σ^2. An unbiased estimator for the mean, μ, based on the sample $\vec{Y}$, is the *sample mean*

$$\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i. \tag{3.97}$$

An unbiased estimator of the variance, σ^2, based on the sample $\vec{Y}$, is the *sample variance*

$$S_Y^2 = \frac{1}{N-1}\sum_{i=1}^{N} [Y_i - \bar{Y}]^2, \tag{3.98}$$

whereas the *sample standard deviation*, S_Y, is just the square root of S_Y^2.

For a continuous-time history, $\vec{X} = \{X_t : A \le t \le B\}$, the *sample time average* is

$$\overline{X} = \frac{1}{B-A}\int_A^B X_t\, dt. \tag{3.99}$$

For a sample of pairs $\vec{Z} = \{(X_1, Y_1), (X_2, Y_2), \ldots, (X_N, Y_N)\}$, with common joint distributions for all pairs (X_j, Y_j), $j = 1, \ldots, N$, a common estimator of the correlation coefficient, $\rho(X, Y)$, is the *sample correlation coefficient*

$$r(X,Y) = \frac{\sum_{i=1}^{N}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{N}(X_i - \overline{X})^2} \times \sqrt{\sum_{i=1}^{N}(Y_i - \overline{Y})^2}}. \tag{3.100}$$

Note that all the estimators above are *point estimators*, that is, they provide a scalar estimate of some unknown parameter. In addition, an estimate of an interval in which the unknown parameter lies can also be established. Specifically, let C be some estimator of an unknown parameter θ. A $(1 - \alpha)$ *confidence interval* for θ is a random interval of the form $[A, B]$, such that $\Pr\{A \le \theta \le B\} = 1 - \alpha$. In other words, a confidence interval $[A,\ B]$ contains the unknown parameter θ with probability $(1 - \alpha)$. Usually, the confidence interval is of the form $[A, B] = [C - D_1, C + D_2]$ for some random offsets, D_1 and D_2.

3.11 HYPOTHESIS TESTING

Hypothesis testing is statistical decision making. The modeler formulates two complementary hypotheses, called the *null hypothesis* (denoted by H_0) and the *alternative hypothesis* (denoted by H_1). A decision is traditionally applied to the null hypothesis, which is either *accepted* or *rejected*. Consequently, two types of errors are possible:

Type I: Rejecting H_0 erroneously

Type II: Accepting H_0 erroneously

The goal of hypothesis testing is to reject (or accept) H_0, such that if H_0 is in fact true, then the probability of erroneously rejecting it (type I error) does not exceed some prescribed probability, α, called the *confidence level* or *significance level*. The smaller is α, the higher is the confidence in a corresponding rejection decision.

For example, suppose we wish to compare the failure rates δ_1 and δ_2 of machines 1 and 2, respectively, at significance level $\alpha = 0.05$. The associated hypotheses follow:

$$\begin{cases} H_0: & \delta_1 \leq \delta_2 \\ H_1: & \delta_1 > \delta_2. \end{cases}$$

The modeler then forms a *test statistic*, T, from some observed sample data with a known distribution under the null hypothesis, H_0. In our case, the statistic might be the difference of two failure rate estimates based on some failure data. The state space S of T is then partitioned into two disjoint regions $S = R_0 \cup R_1$, where R_0 is the *acceptance region*, and R_1 is the *rejection region*, such that the probability of type I error does not exceed, say, $\alpha = 0.05$. In practice, the analyst computes a realization t of T and decides to accept or reject H_0, according as t fell in region R_0 or R_1, respectively. A *critical value*, c, which depends on the significance level and the test statistic, is often used to separate the acceptance and rejection regions. We have noted that intervals comprising individual regions are often constructed as confidence intervals at the corresponding confidence levels.

An alternative approach to hypothesis testing is to compute the *probability value* (commonly abbreviated to *p-value*) of the realization of the test statistic, t, where p is the smallest significance level,

$$\alpha_{\min} = p, \tag{3.101}$$

for which the computed test statistic, t, can be rejected (often p is computed as $p = \Pr\{T > t\}$). To understand this concept, note that in hypothesis testing, we first fix α (and therefore the critical values that define the acceptance and rejection regions), and then decide whether to accept or reject, depending on the region in which the test statistic t fell. We reject the null hypothesis when $p \leq \alpha$, and accept it when $p > \alpha$. Conversely, suppose we do not fix α before computing the test statistic t, but allow it to "float." We would like to know how small can α be made and still permit the null hypothesis to be rejected. This means that we seek the smallest α that satisfies $p \leq \alpha$. Clearly, the requisite $\alpha = \alpha_{\min}$ is given by Eq. 3.101.

The p-value contains a considerable amount of information on the *quality* of our test decision. Not only can we decide whether to reject the null hypothesis or accept it, but we can also obtain an idea on how "strongly" we reject or accept it. The smaller p is compared to α, the stronger is its rejection; conversely, the larger p is relative to α, the stronger is its acceptance. For this reason the p-value is also called the *observed level* of the test. For more information on statistical issues in estimation and hypothesis testing, see Hoel et al. (1971b) and Devore (1991).

EXERCISES

1. Let X be a random variable uniformly distributed over the interval [2, 6], and let Y be a random variable distributed according to Tria(2, 3, 7).