CHAPTER 24

Probability Theory

In decision-making problems, one is often faced with making decisions based upon phenomena that have uncertainty associated with them. This uncertainty is caused by inherent variation due to sources of variation that elude control or the inconsistency of natural phenomena. Rather than treat this variability qualitatively, one can incorporate it into the mathematical model and thus handle it quantitatively. This generally can be accomplished if the natural phenomena exhibit some degree of regularity, so that their variation can be described by a probability model. The ensuing sections are concerned with methods for characterizing these probability models.

24.1 SAMPLE SPACE

Suppose the demand for a product over a period of time, say a month, is of interest. From a realistic point of view, demand is not generally constant but exhibits the type of variation alluded to in the introduction. Suppose an experiment that will result in observing the demand for the product during the month is run. Whereas the outcome of the experiment cannot be predicted exactly, each *possible* outcome can be described. The demand during the period can be any one of the values 0, 1, 2, . . . , that is, the entire set of nonnegative integers. The set of all possible outcomes of the experiment is called the **sample space** and will be denoted by Ω. Each outcome in the sample space is called a *point* and will be denoted by ω. Actually, in the experiment just described, the possible demands may be bounded from above by N, where N would represent the size of the population that has any use for the product. Hence, the sample space would then consist of the set of the integers 0, 1, 2, . . . , N. Strictly speaking, the sample space is much more complex than just described. In fact, it may be extremely difficult to characterize precisely. Associated with this experiment are such factors as the dates and times that the demands occur, the prevailing weather, the disposition of the personnel meeting the demand, and so on. Many more factors could be listed, most of which are irrelevant. Fortunately, as noted in the next section, it is not necessary to describe completely the sample space, but only to record those factors that appear to be necessary for the purpose of the experiment.

Another experiment may be concerned with the time until the first customer arrives at a store. Since the first customer may arrive at any time until the store closes (assuming an 8-hour day), for the purpose of this experiment, the sample space can be considered to be all

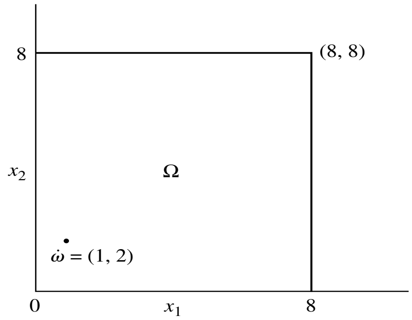


FIGURE 24.1
The sample space of the arrival time experiment over two days.

points on the real line between zero and 8 hours. Thus, Ω consists of all points ω such that

$$0 \leq \omega \leq 8.\dagger$$

Now consider a third example. Suppose that a modification of the first experiment is made by observing the demands during the first 2 months. The sample space Ω consists of all points (x_1,x_2), where x_1 represents the demand during the first month, $x_1 = 0, 1, 2, \ldots$, and x_2 represents the demand during the second month, $x_2 = 0, 1, 2, \ldots$. Thus, Ω consists of the set of all possible points ω, where ω represents a pair of nonnegative integer values (x_1,x_2). The point $\omega = (3,6)$ represents a possible outcome of the experiment where the demand in the first month is 3 units and the demand in the second month is 6 units. In a similar manner, the experiment can be extended to observing the demands during the first n months. In this situation Ω consists of all possible points $\omega = (x_1, x_2, \ldots, x_n)$, where x_i represents the demand during the ith month.

The experiment that is concerned with the time until the first arrival appears can also be modified. Suppose an experiment that measures the times of the arrival of the first customer on each of 2 days is performed. The set of all possible outcomes of the experiment Ω consists of all points (x_1,x_2), $0 \leq x_1, x_2 \leq 8$, where x_1 represents the time the first customer arrives on the first day, and x_2 represents the time the first customer arrives on the second day. Thus, Ω consists of the set of all possible points ω, where ω represents a point in two space lying in the square shown in Fig. 24.1.

This experiment can also be extended to observing the times of the arrival of the first customer on each of n days. The sample space Ω consists of all points $\omega = (x_1, x_2, \ldots, x_n)$, such that $0 \leq x_i \leq 8$ $(i = 1, 2, \ldots, n)$, where x_i represents the time the first customer arrives on the ith day.

An **event** is defined as a set of outcomes of the experiment. Thus, there are many events that can be of interest. For example, in the experiment concerned with observing the demand for a product in a given month, the set $\{\omega = 0, \omega = 1, \omega = 2, \ldots, \omega = 10\}$ is the event that the demand for the product does not exceed 10 units. Similarly, the set $\{\omega = 0\}$ denotes the event of no demand for the product during the month. In the experiment which measures the times of the arrival of the first customer on each of 2 days, the set $\{\omega = (x_1, x_2); x_1 < 1, x_2 < 1\}$ is the event that the first arrival on each day occurs before the first hour. It is evident that any subset of the sample space, e.g., any point, collection of points, or the entire sample space, is an event.

Events may be combined, thereby resulting in the formation of new events. For any two events E_1 and E_2, the new event $E_1 \cup E_2$, referred to as the *union* of E_1 and E_2, is

†It is assumed that at least one customer arrives each day.

defined to contain all points in the sample space that are in either E_1 or E_2, or in both E_1 and E_2. Thus, the event $E_1 \cup E_2$ will occur if either E_1 or E_2 occurs. For example, in the demand experiment, let E_1 be the event of a demand in a single month of zero or 1 unit, and let E_2 be the event of a demand in a single month of 1 or 2 units. The event $E_1 \cup E_2$ is just $\{\omega = 0, \omega = 1, \omega = 2\}$, which is just the event of a demand of 0, 1, or 2 units.

The intersection of two events E_1 and E_2 is denoted by $E_1 \cap E_2$ (or equivalently by E_1E_2). This new event $E_1 \cap E_2$ is defined to contain all points in the sample space that are in both E_1 and E_2. Thus, the event $E_1 \cap E_2$ will occur only if both E_1 and E_2 occur. In the aforementioned example, the event $E_1 \cap E_2$ is $\{\omega = 1\}$, which is just the event of a demand of 1 unit.

Finally, the events E_1 and E_2 are said to be *mutually exclusive* (or disjoint) if their intersection does not contain any points. In the example, E_1 and E_2 are not disjoint. However, if the event E_3 is defined to be the event of a demand of 2 or 3 units, then $E_1 \cap E_3$ is disjoint. Events that do not contain any points, and therefore cannot occur, are called *null events.* (Or course, all these definitions can be extended to any finite number of events.)

24.2 RANDOM VARIABLES

It may occur frequently that in performing an experiment one is not interested directly in the entire sample space or in events defined over the sample space. For example, suppose that the experiment which measures the times of the first arrival on 2 days was performed to determine at what time to open the store. Prior to performing the experiment, the store owner decides that if the average of the arrival times is greater than an hour, thereafter he will not open his store until 10 A.M. (9 A.M. being the previous opening time). The average of x_1 and x_2 (the two arrival times) is not a point in the sample space, and hence he cannot make his decision by looking directly at the outcome of his experiment. Instead, he makes his decision according to the results of a rule that assigns the average of x_1 and x_2 to *each point* (x_1,x_2) in Ω. This resultant set is then partitioned into two parts: those points below 1 and those above 1. If the observed result of this rule (average of the two arrival times) lies in the partition with points greater than 1, the store will be opened at 10 A.M.; otherwise, the store will continue to open at 9 A.M. The rule that assigns the average of x_1 and x_2 to each point in the sample space is called a random variable. Thus, a **random variable** is a numerically valued function defined over the sample space. Note that a function is, in a mathematical sense, just a rule that assigns a number to each value in the domain of definition, in this context the sample space.

Random variables play an extremely important role in probability theory. Experiments are usually very complex and contain information that may or may not be superfluous. For example, in measuring the arrival time of the first customer, the color of his shoes may be pertinent. Although this is unlikely, the prevailing weather may certainly be relevant. Hence, the choice of the random variable enables the experimenter to describe the factors of importance to him and permits him to discard the superfluous characteristics that may be extremely difficult to characterize.

There is a multitude of random variables associated with each experiment. In the experiment concerning the arrival of the first customer on each of 2 days, it has been pointed out already that the average of the arrival times $\overline{X}$ is a random variable. Notationally, random variables will be characterized by capital letters, and the values the random variable takes on will be denoted by lowercase letters. Actually, to be precise, $\overline{X}$ should be written as $\overline{X}(\omega)$, where ω is any point shown in the square in Fig. 24.1 because $\overline{X}$ is a function. Thus, $\overline{X}(1,2) = (1 + 2)/2 = 1.5$, $\overline{X}(1.6,1.8) = (1.6 + 1.8)/2 = 1.7$, $\overline{X}(1.5,1.5) = (1.5 + 1.5)/2 = 1.5$, $\overline{X}(8,8) = (8 + 8)/2 = 8$. The values that the random variable $\overline{X}$ takes on are the set of values $\bar{x}$ such that $0 \le \bar{x} \le 8$. Another random variable, X_1, can be described as follows: For each ω in Ω, the random variable (numerically valued function) disregards the x_2 coordinate and transforms the x_1 coordinate into itself. This random variable, then, represents the arrival time of the first customer on the first day. Hence, $X_1(1,2) = 1$, $X_1(1.6,1.8) = 1.6$, $X_1(1.5,1.5) = 1.5$, $X_1(8,8) = 8$. The values the random variable X_1 takes on are the set of values x_1 such that $0 \le x_1 \le 8$. In a similar manner, the random variable X_2 can be described as representing the arrival time of the first customer on the second day. A third random variable, S^2, can be described as follows: For each ω in Ω, the random variable computes the sum of squares of the deviations of the coordinates about their average; that is, $S^2(\omega) = S^2(x_1, x_2) = (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2$. Hence, $S^2(1,2) = (1 - 1.5)^2 + (2 - 1.5)^2 = 0.5$, $S^2(1.6,1.8) = (1.6 - 1.7)^2 + (1.8 - 1.7)^2 = 0.02$, $S^2(1.5,1.5) = (1.5 - 1.5)^2 + (1.5 - 1.5)^2 = 0$, $S^2(8,8) = (8 - 8)^2 + (8 - 8)^2 = 0$. It is evident that the values the random variable S^2 takes on are the set of values s^2 such that $0 \le s^2 \le 32$.

All the random variables just described are called *continuous* random variables because they take on a continuum of values. *Discrete* random variables are those that take on a finite or countably infinite set of values.[1] An example of a discrete random variable can be obtained by referring to the experiment dealing with the measurement of demand. Let the discrete random variable X be defined as the demand during the month. (The experiment consists of measuring the demand for 1 month). Thus, $X(0) = 0$, $X(1) = 1$, $X(2) = 2, \ldots,$ so that the random variable takes on the set of values consisting of the integers. Note that Ω and the set of values the random variable takes on are identical, so that this random variable is just the identity function.

From the above paragraphs it is evident that any function of a random variable is itself a random variable because a function of a function is also a function. Thus, in the previous examples $\bar{X} = (X_1 + X_2)/2$ and $S^2 = (X_1 - \bar{X})^2 + (X_2 - \bar{X})^2$ can also be recognized as random variables by noting that they are functions of the random variables X_1 and X_2.

This text is concerned with random variables that are real-valued functions defined over the real line or a subset of the real line.

24.3 PROBABILITY AND PROBABILITY DISTRIBUTIONS

Returning to the example of the demand for a product during a month, note that the actual demand is not a constant; instead, it can be expected to exhibit some "variation." In particular, this variation can be described by introducing the concept of probability defined over events in the sample space. For example, let E be the event $\{\omega = 0, \omega = 1, \omega = 2, \ldots, \omega = 10\}$. Then intuitively one can speak of $P\{E\}$, where $P\{E\}$ is referred to as the probability of having a demand of 10 or less units. Note that $P\{E\}$ can be thought of as a numerical value associated with the event E. If $P\{E\}$ is known for all events E in the sample space, then some "information" is available about the demand that can be expected to occur. Usually these numerical values are difficult to obtain, but nevertheless their existence can be postulated. To define the concept of probability rigorously is beyond the scope of this text. However, for most purposes it is sufficient to postulate the existence of numerical values $P\{E\}$ associated with events E in the sample space. The value

[1]A countably infinite set of values is a set whose elements can be put into one-to-one correspondence with the set of positive integers. The set of odd integers is countably infinite. The 1 can be paired with 1, 3 with 2, 5 with $3, \ldots, 2n - 1$ with n. The set of all real numbers between 0 and $\frac{1}{2}$ is not countably infinite because there are too many numbers in the interval to pair with the integers.

$P\{E\}$ is called the *probability* of the occurrence of the event E. Furthermore, it will be assumed that $P\{E\}$ satisfies the following reasonable properties:

1. $0 \leq P\{E\} \leq 1$**.** This implies that the probability of an event is always nonnegative and can never exceed 1.
2. If E_0 is an event that cannot occur (a null event) in the sample space, then $P\{E_0\} = 0$. Let E_0 denote the event of obtaining a demand of -7 units. Then $P\{E_0\} = 0$.
3. $P\{\Omega\} = 1$. If the event consists of obtaining a demand between 0 and N, that is, the entire sample space, the probability of having some demand between 0 and N is certain.
4. If E_1 and E_2 are disjoint(mutually exclusive) events in Ω, then $P\{E_1 \cup E_2\} = P\{E_1\} + P\{E_2\}$. Thus, if E_1 is the event of 0 or 1, and E_2 is the event of a demand of 4 or 5, then the probability of having a demand of 0, 1, 4, or 5, that is, $\{E_1 \cup E_2\}$, is given by $P\{E_1\} + P\{E_2\}$.

Although these properties are rather formal, they do conform to one's intuitive notion about probability. Nevertheless, these properties cannot be used to obtain values for $P\{E\}$. Occasionally the determination of exact values, or at least approximate values, is desirable. Approximate values, together with an interpretation of probability, can be obtained through a frequency interpretation of probability. This may be stated precisely as follows. Denote by n the number of times an experiment is performed and by m the number of successful occurrences of the event E in the n trials. Then $P\{E\}$ can be interpreted as

$$P\{E\} = \lim_{n \to \infty} \frac{m}{n},$$

assuming the limit exists for such a phenomenon. The ratio m/n can be used to approximate $P\{E\}$. Furthermore, m/n satisfies the properties required of probabilities; that is,

1. $0 \leq m/n \leq 1$.
2. $0/n = 0$. (If the event E cannot occur, then $m = 0$.)
3. $n/n = 1$. (If the event E must occur every time the experiment is performed, then $m = n$.)
4. $(m_1 + m_2)/n = m_1/n + m_2/n$ if E_1 and E_2 are disjoint events. (If the event E_1 occurs m_1 times in the n trials and the event E_2 occurs m_2 times in the n trials, and E_1 and E_2 are disjoint, then the total number of successful occurrences of the event E_1 or E_2 is just $m_1 + m_2$.)

Since these properties are true for a finite n, it is reasonable to expect them to be true for

$$P\{E\} = \lim_{n \to \infty} \frac{m}{n}.$$

The trouble with the frequency interpretation as a definition of probability is that it is not possible to actually determine the probability of an event E because the question "How large must n be?" cannot be answered. Furthermore, such a definition does not permit a logical development of the theory of probability. However, a rigorous definition of probability, or finding methods for determining exact probabilities of events, is not of prime importance here.

The existence of probabilities, defined over events E in the sample space, has been described, and the concept of a random variable has been introduced. Finding the relation between probabilities associated with events in the sample space and "probabilities" associated with random variables is a topic of considerable interest.

Associated with every random variable is a **cumulative distribution function** (**CDF**). To define a CDF it is necessary to introduce some additional notation. Define the symbol $E_b^X = \{\omega | X(\omega) \leq b\}$ (or equivalently, $\{X \leq b\}$) as the set of outcomes ω in the sample space forming the event E_b^X such that the random variable X takes on values less than or equal to b.† Then $P\{E_b^X\}$ is just the probability of this event. Note that this probability is well defined because E_b^X is an event in the sample space, and this event depends upon both the random variable that is of interest and the value of b chosen. For example, suppose the experiment that measures the demand for a product during a month is performed. Let $N = 99$, and assume that the events $\{0\}, \{1\}, \{2\}, \ldots, \{99\}$ each has probability equal to $1/100$; that is, $P\{0\} = P\{1\} = P\{2\} = \cdots = P\{99\} = 1/100$. Let the random variable X be the square of the demand, and choose b equal to 150. Then

$$E_{150}^X = \{\omega \mid X(\omega) \leq 150\} = \{X \leq 150\}$$

is the set $E_{150}^X = \{0,1,2,3,4,5,6,7,8,9,10,11,12\}$ (since the square of each of these numbers is less than 150). Furthermore,

$$P\{E_{150}^X\} = \frac{1}{100} + \frac{1}{100} + \frac{1}{100} + \frac{1}{100} + \frac{1}{100} + \frac{1}{100} + \frac{1}{100} + \frac{1}{100} + \frac{1}{100} + \frac{1}{100} + \frac{1}{100} + \frac{1}{100} + \frac{1}{100} = \frac{13}{100}.$$

Thus, $P\{E_{150}^X\} = P\{X \leq 150\} = 13/100$.

For a given random variable X, $P\{X \leq b\}$, denoted by $F_X(b)$, is called the CDF of the random variable X and is defined for all real values of b. Where there is no ambiguity, the CDF will be denoted by $F(b)$; that is,

$$F(b) = F_X(b) = P\{E_b^X\} = P\{\omega \mid X(\omega) \leq b\} = P\{X \leq b\}.$$

Although $P\{X \leq b\}$ is defined through the event E_b^X in the sample space, it will often be read as the "probability" that the random variable X takes on a value less than or equal to b. The reader should interpret this statement properly, i.e., in terms of the event E_b^X.

As mentioned, each random variable has a cumulative distribution function associated with it. This is not an arbitrary function but is induced by the probabilities associated with events of the form E_b^X defined over the sample space Ω. Furthermore, the CDF of a random variable is a numerically valued function defined for all $b, -\infty \leq b \leq \infty$, having the following properties:

1. $F_X(b)$ is a nondecreasing function of b,
2. $\lim_{b \to -\infty} F_X(b) = F_X(-\infty) = 0$,
3. $\lim_{b \to +\infty} F_X(b) = F_X(+\infty) = 1$.

The CDF is a versatile function. Events of the form

$$\{\omega \mid a < X(\omega) \leq b\},$$

that is, the set of outcomes ω in the sample space such that the random variable X takes on values greater than a but not exceeding b, can be expressed in terms of events of the form E_b^X. In particular, E_b^X can be expressed as the union of two disjoint sets; that is,

$$E_b^X = E_a^X \cup \{\omega \mid a < X(\omega) \leq b\}.$$

Thus, $P\{\omega \mid a < X(\omega) \leq b\} = P\{a < X \leq b\}$ can easily be seen to be

$$F_X(b) - F_X(a).$$

As another example, consider the experiment that measures the times of the arrival of the first customer on each of 2 days. Ω consists of all points (x_1, x_2) such that $0 \leq x_1, x_2 \leq 8$,

†The notation $\{X \leq b\}$ suppresses the fact that this is really an event in the sample space. However, it is simpler to write, and the reader is cautioned to interpret it properly, i.e., as the set of outcomes ω in the sample space, $\{\omega \mid X(\omega) \leq b\}$.

where x_1 represents the time the first customer arrives on the first day, and x_2 represents the time the first customer arrives on the second day. Consider all events associated with this experiment, and assume that the probabilities of such events can be obtained. Suppose $\overline{X}$, the average of the two arrival times, is chosen as the random variable of interest and that $E_b^{\overline{X}}$ is the set of outcomes ω in the sample space forming the event $E_b^{\overline{X}}$ such that $\overline{X} \leq b$. Hence, $F_{\overline{X}}(b) = P\{E_b^{\overline{X}}\} = P\{\overline{X} \leq b\}$. To illustrate how this can be evaluated, suppose that $b = 4$ hours. All the values of x_1, x_2 are sought such that $(x_1 + x_2)/2 \leq 4$ or $x_1 + x_2 \leq 8$. This is shown by the shaded area in Fig. 24.2. Hence, $F_{\overline{X}}(b)$ is just the probability of a successful occurrence of the event given by the shaded area in Fig. 24.2. Presumably $F_{\overline{X}}(b)$ can be evaluated if probabilities of such events in the sample space are known.

Another random variable associated with this experiment is X_1, the time of the arrival of the first customer on the first day. Thus, $F_{X_1}(b) = P\{X_1 \leq b\}$, which can be obtained simply if probabilities of events over the sample space are given.

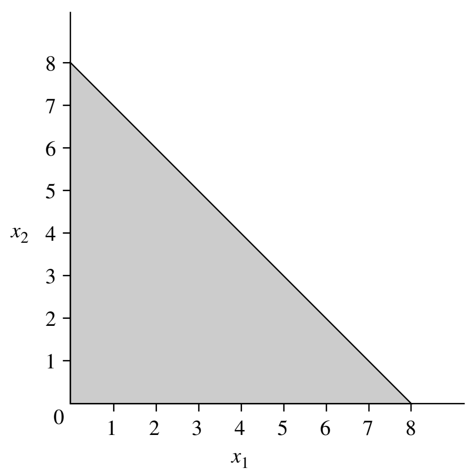
There is a simple frequency interpretation for the cumulative distribution function of a random variable. Suppose an experiment is repeated n times, and the random variable X is observed each time. Denote by $x_1, x_2, \ldots, x_n$ the outcomes of these n trials. Order these outcomes, letting $x_{(1)}$ be the smallest observation, $x_{(2)}$ the second smallest, $\ldots$, $x_{(n)}$ the largest. Plot the following step function $F_n(x)$:

$$
\begin{array}{ll}
\text{For } x < x_{(1)}, & \text{let } F_n(x) = 0. \\
\text{For } x_{(1)} \leq x < x_{(2)}, & \text{let } F_n(x) = \dfrac{1}{n}. \\
\text{For } x_{(2)} \leq x < x_{(3)}, & \text{let } F_n(x) = \dfrac{2}{n}. \\
\vdots & \\
\text{For } x_{(n-1)} \leq x < x_{(n)}, & \text{let } F_n(x) = \dfrac{n-1}{n}. \\
\text{For } x \geq x_{(n)}, & \text{let } F_n(x) = \dfrac{n}{n} = 1.
\end{array}
$$

Such a plot is given in Fig. 24.3 and is seen to "jump" at the values that the random variable takes on.

$F_n(x)$ can be interpreted as the fraction of outcomes of the experiment less than or equal to x and is called the *sample CDF*. It can be shown that as the number of repetitions n of the experiment gets large, the sample CDF approaches the CDF of the random variable X.

FIGURE 24.2
The shaded area represents the event $E_b^{\overline{X}} = \{\overline{X} \leq 4\}$.



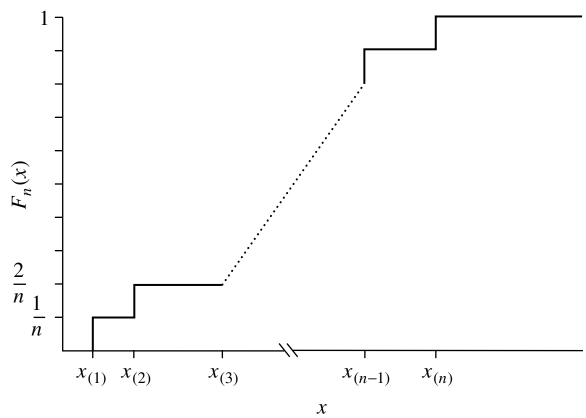


FIGURE 24.3
A sample cumulative distribution function.

In most problems encountered in practice, one is not concerned with events in the sample space and their associated probabilities. Instead, interest is focused on random variables and their associated cumulative distribution functions. Generally, a random variable (or random variables) is chosen, and some assumption is made about the form of the CDF or about the random variable. For example, the random variable X_1, the time of the first arrival on the first day, may be of interest, and an assumption may be made that the form of its CDF is exponential. Similarly, the same assumption about X_2, the time of the first arrival on the second day, may also be made. If these assumptions are valid, then the CDF of the random variable $\overline{X} = (X_1 + X_2)/2$ can be derived. Of course, these assumptions about the form of the CDF are not arbitrary and really imply assumptions about probabilities associated with events in the sample space. Hopefully, they can be substantiated by either empirical evidence or theoretical considerations.

24.4 CONDITIONAL PROBABILITY AND INDEPENDENT EVENTS

Often experiments are performed so that some results are obtained early in time and some later in time. This is the case, for example, when the experiment consists of measuring the demand for a product during each of 2 months; the demand during the first month is observed at the end of the first month. Similarly, the arrival times of the first two customers on each of 2 days are observed sequentially in time. This early information can be useful in making predictions about the subsequent results of the experiment. Such information need not necessarily be associated with time. If the demand for two products during a month is investigated, knowing the demand of one may be useful in assessing the demand for the other. To utilize this information the concept of "conditional probability," defined over events occurring in the sample space, is introduced.

Consider two events in the sample space E_1 and E_2, where E_1 represents the event that has occurred, and E_2 represents the event whose occurrence or nonoccurrence is of interest. Furthermore, assume that $P\{E_1\} > 0$. The **conditional probability** of the occurrence of the event E_2, given that the event E_1 has occurred, $P\{E_2 | E_1\}$, is defined to be

$$P\{E_2 | E_1\} = \frac{P\{E_1 \cap E_2\}}{P\{E_1\}},$$

where $\{E_1 \cap E_2\}$ represents the event consisting of all points ω in the sample space common to *both* E_1 and E_2. For example, consider the experiment that consists of observing the demand for a product over each of 2 months. Suppose the sample space Ω consists of all points $\omega = (x_1,x_2)$, where x_1 represents the demand during the first month, and x_2 represents the demand during the second month, $x_1, x_2 = 0, 1, 2, \ldots, 99$. Furthermore, it is known that the demand during the first month has been 10. Hence, the event E_1, which consists of the points (10,0), (10,1), (10,2), . . . , (10,99), has occurred. Consider the event E_2, which represents a demand for the product in the second month that does not exceed 1 unit. This event consists of the points (0,0), (1,0), (2,0), . . . , (10,0), . . . , (99,0), (0,1), (1,1), (2,1), . . . , (10,1), . . . , (99,1). The event $\{E_1 \cap E_2\}$ consists of the points (10,0) and (10,1). Hence, the probability of a demand which does not exceed 1 unit in the second month, given that a demand of 10 units occurred during the first month, that is, $P\{E_2|E_1\}$, is given by

$$\begin{aligned} P\{E_2|E_1\} &= \frac{P\{E_1 \cap E_2\}}{P\{E_1\}} \\ &= \frac{P\{\omega = (10,0),\, \omega = (10,1)\}}{P\{\omega = (10,0),\, \omega = (10,1), \ldots, \omega = (10,99)\}}. \end{aligned}$$

The definition of conditional probability can be given a frequency interpretation. Denote by n the number of times an experiment is performed, and let n_1 be the number of times the event E_1 has occurred. Let n_{12} be the number of times that the event $\{E_1 \cap E_2\}$ has occurred in the n trials, The ratio n_{12}/n_1 is the proportion of times that the event E_2 occurs when E_1 has also occurred; that is, n_{12}/n_1 is the conditional relative frequency of E_2, given that E_1 has occurred. This relative frequency n_{12}/n_1 is then equivalent to $(n_{12}/n)/(n_1/n)$. Using the frequency interpretation of probability for large n, n_{12}/n is approximately $P\{E_1 \cap E_2\}$, and n_1/n is approximately $P\{E_1\}$, so that the conditional relative frequency of E_2, given E_1, is approximately $P\{E_1 \cap E_2\}/P\{E_1\}$.

In essence, if one is interested in conditional probability, he is working with a reduced sample space, i.e., from Ω to E_1, modifying other events accordingly. Also note that conditional probability has the four properties described in Sec. 24.3; that is,

1. $0 \le P\{E_2|E_1\} \le 1$.
2. If E_2 is an event that cannot occur, then $P\{E_2|E_1\} = 0$.
3. If the event E_2 is the entire sample space Ω, then $P\{E_2|E_1\} = 1$.
4. If E_2 and E_3 are disjoint events in Ω, then

$$P\{(E_2 \cup E_3)|E_1\} = P\{E_2|E_1\} + P\{E_3|E_1\}.$$

In a similar manner, the conditional probability of the occurrence of the event E_1, given that the event E_2 has occurred, can be defined. If $P\{E_2\} > 0$, then

$$P\{E_1|E_2\} = P\{E_1 \cap E_2\}/P\{E_2\}.$$

The concept of conditional probability was introduced so that advantage could be taken of information about the occurrence or nonoccurrence of events. It is conceivable that information about the occurrence of the event E_1 yields no information about the occurrence or nonoccurrence of the event E_2. If $P\{E_2|E_1\} = P\{E_2\}$, or $P\{E_1|E_2\} = P\{E_1\}$, then E_1 and E_2 are said to be **independent events.** It then follows that if E_1 and E_2 are independent and $P\{E_1\} > 0$, then $P\{E_2|E_1\} = P\{E_1 \cap E_2\}/P\{E_1\} = P\{E_2\}$, so that $P\{E_1 \cap E_2\} = P\{E_1\}\, P\{E_2\}$. This can be taken as an alternative definition of independence of the events E_1 and E_2. It is usually difficult to show that events are independent by using the definition of independence. Instead, it is generally simpler to use the information available about the experiment to postulate whether events are independent. This is usually based upon physical considerations. For example, if the demand for a product during a month is "known" *not* to affect the demand in subsequent months, then the events E_1 and E_2 defined previously can be said to be independent, in which case

$$\begin{aligned} P\{E_2 | E_1\} &= \frac{P\{E_1 \cap E_2\}}{P\{E_1\}} \\ &= \frac{P(\omega = (10,0), \omega = (10,1)\}}{P\{\omega = (10,0), \omega = (10,1), \ldots, \omega = (10,99)\}}, \\ &= \frac{P\{E_1\}P\{E_2\}}{P\{E_1\}} = P\{E_2\} \\ &= P\{\omega = (0,0), \omega = (1,0), \ldots, \omega = (99,0), \omega = (0,1), \\ &\qquad \omega = (1,1), \ldots, \omega = (99,1)\}. \end{aligned}$$

The definition of independence can be extended to any number of events. $E_1, E_2, \ldots, E_n$ are said to be independent events if for *every* subset of these events $E_1^*, E_2^*, \ldots, E_k^*$,

$$P\{E_1^* \cap E_2^* \cap \cdots \cap E_k^*\} = P\{E_1^*\}P\{E_2^*\}\cdots P\{E_k^*\}.$$

Intuitively, this implies that knowledge of the occurrence of any of these events has no effect on the probability of occurrence of any other event.

24.5 DISCRETE PROBABILITY DISTRIBUTIONS

It was pointed out in Sec. 24.2 that one is usually concerned with random variables and their associated probability distributions, and discrete random variables are those which take on a finite or countably infinite set of values. Furthermore, Sec. 24.3 indicates that the CDF for a random variable is given by

$$F_X(b) = P\{\omega | X(\omega) \le b\}.$$

For a discrete random variable X, the event $\{\omega | X(\omega) \le b\}$ can be expressed as the union of disjoint sets; that is,

$$\{\omega | X(\omega) \le b\} = \{\omega | X(\omega) = x_1\} \cup \{\omega | X(\omega) = x_2\} \cup \cdots \cup \{\omega | X(\omega) = x_{[b]}\},$$

where $x_{[b]}$ denotes the largest integer value of the x's less than or equal to b. It then follows that for the discrete random variable X, the CDF can be expressed as

$$\begin{aligned} F_X(b) &= P\{\omega | X(\omega) = x_1\} + P\{\omega | X(\omega) = x_2\} + \cdots + P\{\omega | X(\omega) = x_{[b]}\} \\ &= P\{X = x_1\} + P\{X = x_2\} + \cdots + P\{X = x_{[b]}\}. \end{aligned}$$

This last expression can also be expressed as

$$F_X(b) = \sum_{\text{all } k \le b} P\{X = k\},$$

where k is an index that ranges over all the possible x values which the random variable X can take on.

Let $P_X(k)$ for a specific value of k denote the probability $P\{X = k\}$, so that

$$F_X(b) = \sum_{\text{all } k \le b} P_X(k).$$

This $P_X(k)$ for all possible values of k are called the probability distribution of the discrete random variable X. When no ambiguity exists, $P_X(k)$ may be denoted by $P(k)$.

As an example, consider the discrete random variable that represents the demand for a product in a given month. Let $N = 99$. If it is assumed that $P_X(k) = P\{X = k\} = 1/100$

for all $k = 0, 1, \ldots, 99$, then the CDF for this discrete random variable is given in Fig. 24.4. The probability distribution of this discrete random variable is shown in Fig. 24.5. Of course, the heights of the vertical lines in Fig. 24.5 are all equal because $P_X(0) = P_X(1) = P_x(2) = \cdots = P_X(99)$ in this case. For other random variables X, the $P_X(k)$ need not be equal, and hence the vertical lines will not be constant. In fact, all that is required for the $P_X(k)$ to form a probability distribution is that $P_X(k)$ for each k be nonnegative and

$$\sum_{\text{all k}} P_X(k) = 1.$$

There are several important discrete probability distributions used in operations research work. The remainder of this section is devoted to a study of these distributions.

Binomial Distribution

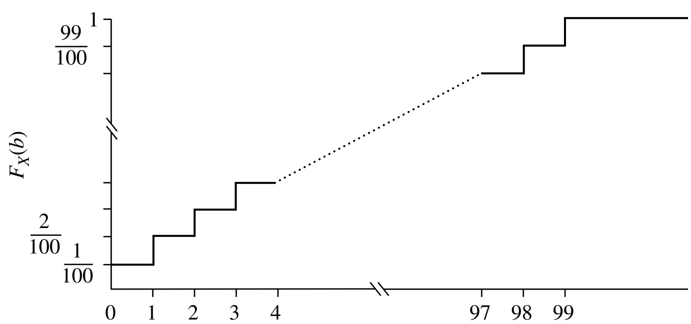
A random variable X is said to have a binomial distribution if its probability distribution can be written as

$$P\{X = k\} = P_X(k) = \frac{n!}{k!(n-k)!} p^k(1-p)^{n-k},$$

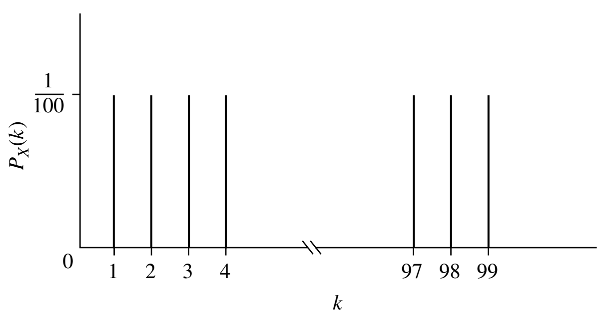
where p is a constant lying between zero and 1, n is any positive integer, and k is also an integer such that $0 \leq k \leq n$. It is evident that $P_x(k)$ is always nonnegative, and it is easily proven that

$$\sum_{k=0}^{n} P_X(k) = 1.$$

■ **FIGURE 24.4**
CDF of the discrete random variable for the example.



■ **FIGURE 24.5**
Probability distribution of the discrete random variable for the example.



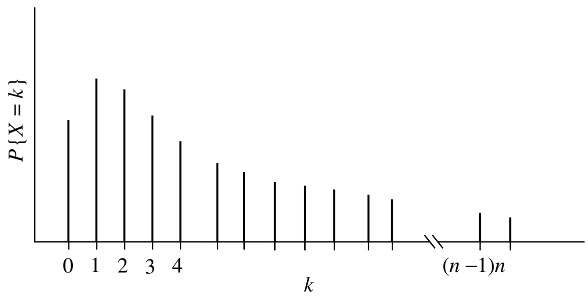


FIGURE 24.6
Binomial distribution with parameters *n* and *p*.

Note that this distribution is a function of the two parameters n and p. The probability distribution of this random variable is shown in Fig. 24.6. An interesting interpretation of the binomial distribution is obtained when $n = 1$:

$$P\{X = 0\} = P_X(0) = 1 - p,$$

and

$$P\{X = 1\} = P_X(1) = p.$$

Such a random variable is said to have a *Bernoulli distribution*. Thus, if a random variable takes on two values, say, 0 or 1, with probability $1 - p$ or p, respectively, a Bernoulli random variable is obtained. The upturned face of a flipped coin is such an example: If a head is denoted by assigning it the number 0 and a tail by assigning it a 1, and if the coin is "fair" (the probability that a head will appear is $\frac{1}{2}$), the upturned face is a Bernoulli random variable with parameter $p = \frac{1}{2}$. Another example of a Bernoulli random variable is the quality of an item. If a defective item is denoted by 1 and a nondefective item by 0, and if p represents the probability of an item being defective, and $1 - p$ represents the probability of an item being nondefective, then the "quality" of an item (defective or nondefective) is a Bernoulli random variable.

If $X_1, X_2, \ldots, X_n$ are independent[1] Bernoulli random variables, each with parameter p, then it can be shown that the random variable

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

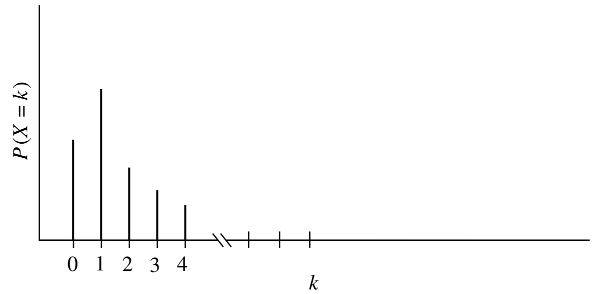
is a binomial random variable with parameters n and p. Thus, if a fair coin is flipped 10 times, with the random variable X denoting the total number of tails (which is equivalent to $X_1 + X_2 + \cdots + X_{10}$), then X has a binomial distribution with parameters 10 and $\frac{1}{2}$; that is,

$$P\{X = k\} = \frac{10!}{k!(10-k)!}\left(\frac{1}{2}\right)^k\left(\frac{1}{2}\right)^{10-k}.$$

Similarly, if the quality characteristics (defective or nondefective) of 50 items are independent Bernoulli random variables with parameter p, the total number of defective items in the 50 sampled, that is, $X = X_1 + X_2 + \cdots + X_{50}$, has a binomial distribution with parameters 50 and p, so that

$$P\{X = k\} = \frac{50!}{k!(50-k)!}\, p^k(1-p)^{50-k}.$$

[1]The concept of independent random variables is introduced in Sec. 24.12. For the present purpose, random variables can be considered independent if their outcomes do not affect the outcomes of the other random variables.



■ **FIGURE 24.7**
Poisson distribution.

Poisson Distribution

A random variable X is said to have a Poisson distribution if its probability distribution can be written as

$$P\{X = k\} = P_X(k) = \frac{\lambda^k e^{-\lambda}}{k!},$$

where λ is a positive constant (the parameter of this distribution), and k is any nonnegative integer. It is evident that $P_X(k)$ is nonnegative, and it is easily shown that

$$\sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} = 1.$$

An example of a probability distribution of a Poisson random variable is shown in Fig. 24.7.

The Poisson distribution is often used in operations research. Heuristically speaking, this distribution is appropriate in many situations where an "event" occurs over a period of time when it is as likely that this "event" will occur in one interval as in any other and the occurrence of an event has no effect on whether or not another occurs. As discussed in Sec. 17.4, the number of customer arrivals in a fixed time is often assumed to have a Poisson distribution. Similarly, the demand for a given product is also often assumed to have this distribution.

Geometric Distribution

A random variable X is said to have a geometric distribution if its probability distribution can be written as

$$P\{X = k\} = P_X(k) = p(1 - p)^{k-1},$$

where the parameter p is a constant lying between 0 and 1, and k takes on the values 1, 2, 3, ... It is clear that $P_X(k)$ is nonnegative, and it is easy to show that

$$\sum_{k=1}^{\infty} p(1 - p)^{k-1} = 1.$$

The geometric distribution is useful in the following situation. Suppose an experiment is performed that leads to a sequence of independent[1] Bernoulli random variables, each with parameter p; that is, $P\{X_1 = 1\} = p$ and $P(X_1 = 0) = 1 - p$, for all i. The random variable X, which is the number of trials occurring until the first Bernoulli random variable takes on the value 1, has a geometric distribution with parameter p.

[1]The concept of independent random variables is introduced in Sec. 24.12. For now, random variables can be considered independent if their outcomes do not affect the outcomes of the other random variables.

24.6 CONTINUOUS PROBABILITY DISTRIBUTIONS

Section 24.2 defined continuous random variables as those random variables that take on a continuum of values. The CDF for a continuous random variable $F_X(b)$ can usually be written as

$$F_X(b) = P\{X(\omega) \le b\} = \int_{-\infty}^{b} f_X(y)dy,$$

where $f_X(y)$ is known as the **density function** of the random variable X. From a notational standpoint, the subscript X is used to indicate the random variable that is under consideration. When there is no ambiguity, this subscript may be deleted, and $f_X(y)$ will be denoted by $f(y)$. It is evident that the CDF can be obtained if the density function is known. Furthermore, a knowledge of the density function enables one to calculate all sorts of probabilities, for example,

$$P\{a < X \le b\} = F(b) - F(a) = \int_{a}^{b} f_X(y)\, dy.$$

Note that strictly speaking the symbol $P\{a < X \le b\}$ relates to the probability that the outcome ω of the experiment belongs to a particular event in the sample space, namely, that event such that $X(\omega)$ is between a and b whenever ω belongs to the event. However, the reference to the event will be suppressed, and the symbol P will be used to refer to the probability that X falls between a and b. It becomes evident from the previous expression for $P\{a < X \le b\}$ that this probability can be evaluated by obtaining the area under the density function between a and b, as illustrated by the shaded area under the density function shown in Fig. 24.8. Finally, if the density function is known, it will be said that the probability distribution of the random variable is determined.

Naturally, the density function can be obtained from the CDF by using the relation

$$\frac{dF_X(y)}{dy} = \frac{d}{dy}\int_{-\infty}^{y} f_X(t)\, dt = f_X(y).$$

For a given value c, $P\{X = c\}$ has not been defined in terms of the density function. However, because probability has been interpreted as an area under the density function, $P\{X = c\}$ will be taken to be zero for all values of c. Having $P\{X = c\} = 0$ does not mean that the appropriate event E in the sample space (E contains those ω such that $X(\omega) = c$) is an impossible event. Rather, the event E can occur, but it occurs with *probability zero.* Since X is a continuous random variable, it takes on a continuum of possible values, so that selecting correctly the actual outcome before experimentation would be rather startling. Nevertheless, some outcome is obtained, so that it is not unreasonable to assume that the preselected outcome has probability zero of occurring. It then follows from $P\{X = c\}$ being equal to zero for all values c that for continuous random variables, and any a and b,

$$P\{a \le X \le b\} = P\{a < X \le b\} = P\{a \le X < b\} = P\{a < X < b\}.$$

Of course, this is not true for discrete random variables.

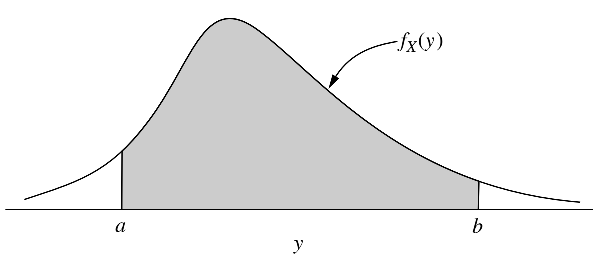


FIGURE 24.8
An example of a density function of a random variable.

In defining the CDF for continuous random variables, it was implied that $f_X(y)$ was defined for values of y from minus infinity to plus infinity because

$$F_X(b) = \int_{-\infty}^{b} f_X(y)\, dy.$$

This causes no difficulty, even for random variables that cannot take on negative values (e.g., the arrival time of the first customer) or are restricted to other regions, because $f_X(y)$ can be defined to be zero over the inadmissible segment of the real line. In fact, the only requirements of a density function are that

1. $f_X(y)$ be nonnegative,

2. $\int_{-\infty}^{\infty} f_X(y)\, dy = 1.$

It has already been pointed out that $f_X(y)$ cannot be interpreted as $P\{X = y\}$ because this probability is always zero. However, $f_X(y)\, dy$ can be interpreted as the probability that the random variable X lies in the infinitesimal interval $(y, y + dy)$, so that, loosely speaking, $f_X(y)$ is a measure of the frequency with which the random variable will fall into a "small" interval near y.

There are several important continuous probability distributions that are used in operations research work. The remainder of this section is devoted to a study of these distributions.

The Exponential Distribution

As was discussed in Sec. 17.4, a continuous random variable whose density is given by

$$f_X(y) = \begin{cases} \dfrac{1}{\theta} e^{-y/\theta}, & \text{for } y \geq 0 \\ 0, & \text{for } y < 0 \end{cases}$$

is known as an exponentially distributed random variable. The exponential distribution is a function of the single parameter θ, where θ is any positive constant. (In Sec. 17.4, we used $\alpha = 1/\theta$ as the parameter instead, but it will be convenient to use θ as the parameter in this chapter.) $f_X(y)$ is a density function because it is nonnegative and integrates to 1; that is,

$$\int_{-\infty}^{\infty} f_X(y)\, dy = \int_{0}^{\infty} \frac{1}{\theta}\, e^{-y/\theta}\, dy = -e^{-y/\theta}\Big|_0^{\infty} = 1.$$

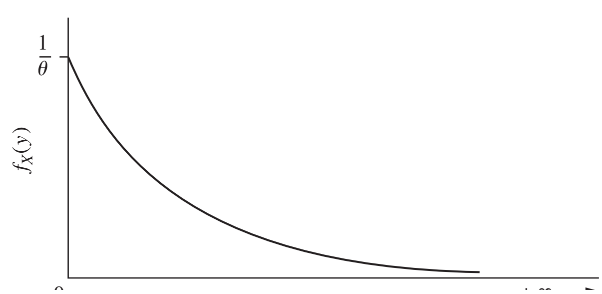
The exponential density function is shown in Fig. 24.9.

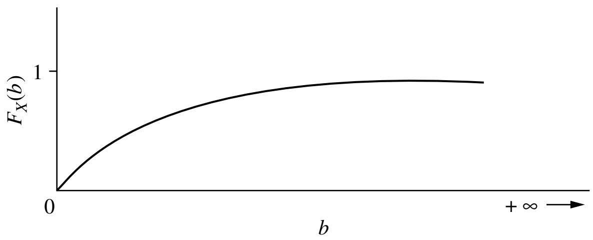
The CDF of an exponentially distributed random variable $f_X(b)$ is given by

$$\begin{aligned} F_X(b) &= \int_{-\infty}^{b} f_X(y)\, dy \\ &= \begin{cases} 0, & \text{for } b < 0 \\ \displaystyle\int_0^b \frac{1}{\theta} e^{-y/\theta} dy = 1 - e^{-b/\theta}, & \text{for } b \geq 0, \end{cases} \end{aligned}$$

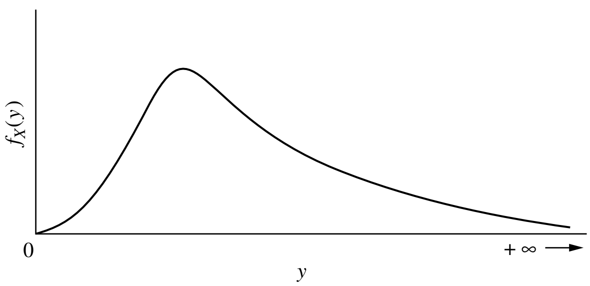
and is shown in Fig. 24.10.

FIGURE 24.9
Density function of the exponential distribution.





■ **FIGURE 24.10**
CDF of the exponential distribution.



■ **FIGURE 24.11**
Gamma density function.

The exponential distribution has had widespread use in operations research. The time between customer arrivals, the length of time of telephone conversations, and the life of electronic components are often assumed to have an exponential distribution. Such an assumption has the important implication that the random variable does not "age." For example, suppose that the life of a vacuum tube is assumed to have an exponential distribution. If the tube has lasted 1,000 hours, the probability of lasting an additional 50 hours is the same as the probability of lasting an additional 50 hours, given that the tube has lasted 2,000 hours. In other words, a brand new tube is no "better" than one that has lasted 1,000 hours. This implication of the exponential distribution is quite important and is often overlooked in practice.

The Gamma Distribution

A continuous random variable whose density is given by

$$f_X(y) = \begin{cases} \dfrac{1}{\Gamma(\alpha)\beta^{\alpha}} y^{(\alpha-1)} e^{-y/\beta}, & \text{for } y \geq 0 \\ 0, & \text{for } y < 0 \end{cases}$$

is known as a gamma-distributed random variable. This density is a function of the two parameters α and β, both of which are positive constants. $\Gamma(\alpha)$ is defined as

$$\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} e^{-t}\, dt, \text{ for all } \alpha > 0.$$

If α is an integer, then repeated integration by parts yields

$$\Gamma(\alpha) = (\alpha - 1)! = (\alpha - 1)(\alpha - 2)(\alpha - 3) \cdots 3 \cdot 2 \cdot 1.$$

With α an integer, the gamma distribution is known in queueing theory as the *Erlang distribution* (as discussed in Sec. 17.7), in which case α is referred to as the *shape parameter.*

A graph of a typical gamma density function is given in Fig. 24.11.

A random variable having a gamma density is useful in its own right as a mathematical representation of physical phenomena, or it may arise as follows: Suppose a customer's service time has an exponential distribution with parameter θ. The random variable T, the total time to service n (independent) customers, has a gamma distribution with parameters n and θ (replacing α and β, respectively); that is,

$$P\{T < t\} = \int_0^t \frac{1}{\Gamma(n)\theta^n} y^{(n-1)} e^{-y/\theta} \, dy.$$

Note that when $n = 1$ (or $\alpha = 1$) the gamma density becomes the density function of an exponential random variable. Thus, sums of independent, exponentially distributed random variables have a gamma distribution.

Another important distribution, the **chi square,** is related to the gamma distribution. If X is a random variable having a gamma distribution with parameters $\beta = 1$ and $\alpha = v/2$ (v is a positive integer), then a new random variable $Z = 2X$ is said to have a chi-square distribution with v degrees of freedom. The expression for the density function is given in Table 24.1 at the end of Sec. 24.8.

The Beta Distribution

A continuous random variable whose density function is given by

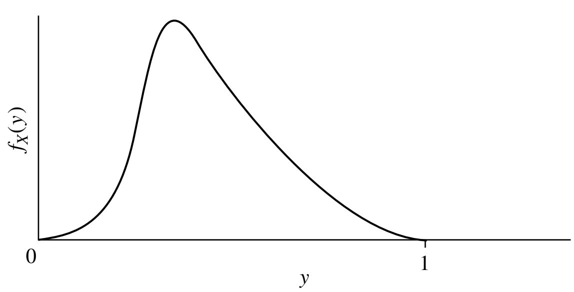
$$f_X(y) = \begin{cases} \dfrac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{(\alpha-1)}(1 - y)^{(\beta-1)}, & \text{for } 0 \le y \le 1 \\ 0, & \text{elsewhere} \end{cases}$$

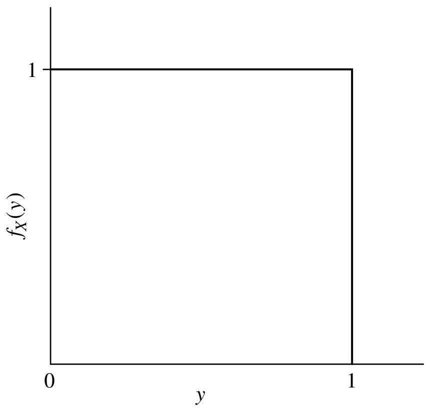
is known as a beta-distributed random variable. This density is a function of the two parameters α and β, both of which are positive constants. A graph of a typical beta density function is given in Fig. 24.12.

Beta distributions form a useful class of distributions when a random variable is restricted to the unit interval. In particular, when $\alpha = \beta = 1$, the beta distribution is called the **uniform distribution** over the unit interval. Its density function is shown in Fig. 24.13, and it can be interpreted as having all the values between zero and 1 equally likely to occur. The CDF for this random variable is given by

$$F_X(b) = \begin{cases} 0, & \text{for } b < 0 \\ b, & \text{for } 0 \le b \le 1 \\ 1, & \text{for } b > 1. \end{cases}$$

FIGURE 24.12
Beta density function.





■ **FIGURE 24.13**
Uniform distribution over the unit interval.

If the density function is to be constant over some other interval, such as the interval $[c, d]$, a uniform distribution over this interval can also be obtained.[1] The density function is given by

$$f_X(y) = \begin{cases} \dfrac{1}{d-c}, & \text{for } c \le y \le d \\ 0, & \text{otherwise.} \end{cases}$$

Although such a random variable is said to have a uniform distribution over the interval $[c, d]$, it is no longer a special case of the beta distribution.

Another important distribution, **Students *t*,** is related to the beta distribution. If X is a random variable having a beta distribution with parameters $\alpha = \mathbf{1/2}$ and $\beta = v/2$ (v is a positive integer), then a new random variable $Z = \sqrt{vX/(1-X)}$ is said to have a Students t (or t) distribution with v degrees of freedom. The percentage points of the t distribution are given in Table 27.6. (Percentage points of the distribution of a random variable Z are the values z_α such that

$$P\{Z > z_\alpha\} = \alpha,$$

where z_α is said to be the 100α percentage point of the distribution of the random variable Z.)

A final distribution related to the beta distribution is the ***F* distribution.** If X is a random variable having a beta distribution with parameters $\alpha = v_1/2$ and $\beta = v_2/2$ (v_1 and v_2 are positive integers), then a new random variable $Z = v_2 X/v_1(1 - X)$ is said to have an F distribution with v_1 and v_2 degrees of freedom.

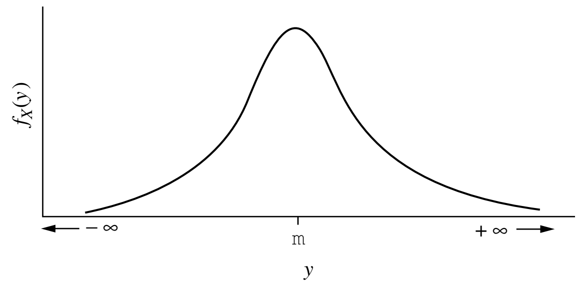
The Normal Distribution

One of the most important distributions in operations research is the normal distribution. A continuous random variable whose density function is given by

$$f_X(y) = \frac{1}{\sqrt{2\pi}\sigma}\, e^{-(y-\mu)^2/2\sigma^2}, \qquad \text{for } -\infty < y < \infty$$

is known as a normally distributed random variable. The density is a function of the two parameters μ and σ, where μ is any constant, and σ is positive. A graph of a typical normal density function is given in Fig. 24.14. This density function is a bell-shaped curve that is

[1]The beta distribution can also be generalized by defining the density function over some fixed interval other than the unit interval.



■ **FIGURE 24.14**
Normal density function.

symmetric around μ. The CDF for a normally distributed random variable is given by

$$F_X(b) = \int_{-\infty}^{b} \frac{1}{\sqrt{2\pi}\sigma} e^{-(y-\mu)^2/2\sigma^2}\, dy.$$

By making the transformation $z = (y - \mu)/\sigma$, the CDF can be written as

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

Hence, although this function is not integrable, it is easily tabulated. Table A5.1 presented in Appendix 5 is a tabulation of

$$\alpha = \int_{K_\alpha}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz$$

as a function of K_α. Hence, to find $F_X(b)$ (and any probability derived from it), Table A5.1 is entered with $K_\alpha = (b - \mu)/\sigma$, and

$$\alpha = \int_{K_\alpha}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz$$

is read from it. $F_X(b)$ is then just $1 - \alpha$. Thus, if $P\{14 < X \leq 18\} = F_X(18) - F_X(14)$ is desired, where X has a normal distribution with $\mu = 10$ and $\sigma = 4$, Table A5.1 is entered with $(18 - 10)/4 = 2$, and $1 - F_X(18) = 0.0228$ is obtained. The table is then entered with $(14 - 10)/4 = 1$, and $1 - F_X(14) = 0.1587$ is read. From these figures, $F_X(18) - F_X(14) = 0.1359$ is found. If K_α is negative, use can be made of the symmetry of the normal distribution because

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

In this case $-(b - \mu)/\sigma$ is positive, and $F_X(b) = \alpha$ is thereby read from the table by entering it with $-(b - \mu)/\sigma$. Thus, suppose it is desired to evaluate the expression

$$P\{2 < X \leq 18\} = F_X(18) - F_X(2).$$

$F_X(18)$ has already been shown to be equal to $1 - 0.0228 = 0.9772$. To find $F_X(2)$ it is first noted that $(2 - 10)/4 = -2$ is negative. Hence, Table A5.1 is entered with $K_\alpha = +2$, and $F_X(2) = 0.0228$ is obtained. Thus,

$$F_X(18) - F_X(2) = 0.9772 - 0.0228 = 0.9544.$$

As indicated previously, the normal distribution is a very important one. In particular, it can be shown that if $X_1, X_2, \ldots, X_n$ are independent,[1] normally distributed random

[1]The concept of independent random variables is introduced in Sec. 24.12. For now, random variables can be considered independent if their outcomes do not affect the outcomes of the other random variables.

variables with parameters $(\mu_1, \sigma_1), (\mu_2, \sigma_2), \ldots, (\mu_n, \sigma_n)$, respectively, then $X = X_1 + X_2 + \cdots + X_n$ is also a normally distributed random variable with parameters

$$\sum_{i=1}^{n} \mu_i$$

and

$$\sqrt{\sum_{i=1}^{n} \sigma_i^2}.$$

In fact, even if $X_1, X_2, \ldots, X_n$ are not normal, then under very weak conditions

$$X = \sum_{i=1}^{n} X_i$$

tends to be normally distributed as n gets large. This is discussed further in Sec. 24.14.

Finally, if C is any constant and X is normal with parameters μ and σ, then the random variable CX is also normal with parameters $C\mu$ and $C\sigma$. Hence, it follows that if $X_1, X_2, \ldots, X_n$ are independent, normally distributed random variables, each with parameters μ and σ, the random variable

$$\overline{X} = \sum_{i=1}^{n} \frac{X_i}{n}$$

is also normally distributed with parameters μ and $\sigma/\sqrt{n}$.

24.7 EXPECTATION

Although knowledge of the probability distribution of a random variable enables one to make all sorts of probability statements, a single value that may characterize the random variable and its probability distribution is often desirable. Such a quantity is the *expected value* of the random variable. One may speak of the expected value of the demand for a product or the expected value of the time of the first customer arrival. In the experiment where the arrival time of the first customer on two successive days was measured, the expected value of the average arrival time of the first customers on two successive days may be of interest.

Formally, the expected value of a random variable X is denoted by $E(X)$ and is given by

$$E(X) = \begin{cases} \displaystyle\sum_{\text{all } k} kP\{\mathrm{X} = k\} = \sum_{\text{all } k} kP_X(k), & \text{if } X \text{ is a discrete random variable} \\ \displaystyle\int_{-\infty}^{\infty} y\, f_X(y)\, dy, & \text{if } X \text{ is a continuous random variable.} \end{cases}$$

For a discrete random variable it is seen that $E(X)$ is just the sum of the products of the possible values the random variable X takes on and their respective associated probabilities. In the example of the demand for a product, where $k = 0, 1, 2, \ldots, 98, 99$ and $P_X(k) = 1/100$ for all k, the expected value of the demand is

$$E(X) = \sum_{k=0}^{99} kP_X(k) = \sum_{k=0}^{99} k\,\frac{1}{100} = 49.5.$$

Note that $E(X)$ need not be a value that the random variable can take on.

If X is a binomial random variable with parameters n and p, the expected value of X is given by

$$E(X) = \sum_{k=0}^{n} k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

and can be shown to equal np.

If the random variable X has a Poisson distribution with parameter λ,

$$E(X) = \sum_{k=0}^{\infty} k \frac{\lambda^k e^{-\lambda}}{k!}$$

and can be shown to equal λ.

Finally, if the random variable X has a geometric distribution with parameter p,

$$E(X) = \sum_{k=1}^{\infty} kp(1-p)^{k-1}$$

and can be shown to equal $1/p$.

For continuous random variables, the expected value can also be obtained easily. If X has an exponential distribution with parameter θ, the expected value is given by

$$E(X) = \int_{-\infty}^{\infty} y f_X(y)\, dy = \int_0^{\infty} y \frac{1}{\theta} e^{-y/\theta}\, dy.$$

This integral is easily evaluated to be

$$E(X) = \theta.$$

If the random variable X has a gamma distribution with parameter α and β the expected value of X is given by

$$\int_{-\infty}^{\infty} y f_X(y)\, dy = \int_0^{\infty} y \frac{1}{\Gamma(\alpha)\beta^{\alpha}} y^{(\alpha-1)} e^{-y/\beta}\, dy = \alpha\beta.$$

If the random variable X has a beta distribution with parameters α and β, the expected value of X is given by

$$\int_{-\infty}^{\infty} y f_X(y)\, dy = \int_0^{1} y \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{(\alpha-1)}(1-y)^{(\beta-1)}\, dy = \frac{\alpha}{\alpha+\beta}.$$

Finally, if the random variable X has a normal distribution with parameters μ and σ, the expected value of X is given by

$$\int_{-\infty}^{\infty} y f_X(y)\, dy = \int_{-\infty}^{\infty} y \frac{1}{\sqrt{2\pi}\sigma} e^{-(y-\mu)^2/2\sigma^2}\, dy = \mu.$$

The expectation of a random variable is quite useful in that it not only provides some characterization of the distribution, but it also has meaning in terms of the average of a sample. In particular, if a random variable is observed again and again and the arithmetic mean $\bar{X}$ is computed, then $\bar{X}$ tends to the expectation of the random variable X as the number of trials becomes large. A precise statement of this property is given in Sec. 24.13. Thus, if the demand for a product takes on the values $k = 0, 1, 2, \ldots, 98, 99$, each with $P_X(k) = 1/100$ for all k, and if demands of $x_1, x_2, \ldots, x_n$ are observed on successive days, then the average of these values, $(x_1 + x_2 + \cdots + x_n)/n$, should be close to $E(X) = 49.5$ if n is sufficiently large.

It is not necessary to confine the discussion of expectation to discussion of the expectation of a random variable X. If Z is some function of X, say, $Z = g(X)$, then $g(X)$ is also a random variable. The expectation of $g(X)$ can be defined as

$$E[g(X)] = \begin{cases} \sum_{\text{all } k} g(k)P\{X = k\} = \sum_{\text{all } k} g(k)P_X(k), & \text{if } X \text{ is a discrete random variable} \\ \int_{-\infty}^{\infty} g(y)\, f_X(y)\, dy, & \text{if } X \text{ is a continuous random variable.} \end{cases}$$

An interesting theorem, known as the "theorem of the unconscious statistician,"[1] states that if X is a continuous random variable having density $f_X(y)$ and $Z = g(X)$ is a function of X having density $h_Z(y)$, then

$$E(Z) = \int_{-\infty}^{\infty} y h_Z(y)\, dy = \int_{-\infty}^{\infty} g(y) f_X(y)\, dy.$$

Thus, the expectation of Z can be found by using its definition in terms of the density of Z or, alternatively, by using its definition as the expectation of a function of X with respect to the density function of X. The identical theorem is true for discrete random variables.

24.8 MOMENTS

If the function g described in the preceding section is given by

$$Z = g(X) = X^j,$$

where j is a positive integer, then the expectation of X^j is called the jth moment about the origin of the random variable X and is given by

$$E(X^j) = \begin{cases} \sum_{\text{all } k} k^j P_X(k), & \text{if } X \text{ is a discrete random variable} \\ \int_{-\infty}^{\infty} y^j f_X(y)\, dy, & \text{if } X \text{ is a continuous random variable.} \end{cases}$$

Note that when $j = 1$ the first moment coincides with the expectation of X. This is usually denoted by the symbol μ and is often called the **mean** or average of the distribution.

Using the theorem of the unconscious statistician, the expectation of $Z = g(X) = CX$ can easily be found, where C is a constant. If X is a continuous random variable, then

$$E(CX) = \int_{-\infty}^{\infty} Cy f_X(y)\, dy = C \int_{-\infty}^{\infty} y f_X(y)\, dy = CE(X).$$

Thus, the expectation of a constant times a random variable is just the constant times the expectation of the random variable. This is also true for discrete random variables.

If the function g described in the preceding section is given by $Z = g(X) = (X - E(X))^j = (X - \mu)^j$, where j is a positive integer, then the expectation of $(X - \mu)^j$ is called the ***j*th moment about the mean** of the random variable X and is given by

$$E(X - E(X))^j = E(X - \mu)^j = \begin{cases} \sum_{\text{all } k} (k - \mu)^j P_X(k), & \text{if } X \text{ is a discrete random variable} \\ \int_{-\infty}^{\infty} (y - \mu)^j f_X(y)\, dy, & \text{if } X \text{ is a continuous random variable.} \end{cases}$$

Note that if $j = 1$, then $E(X - \mu) = 0$. If $j = 2$, then $E(X - \mu)^2$ is called the **variance** of the random variable X and is often denoted by σ^2. The square root of the variance σ

[1]The name for this theorem is motivated by the fact that a statistician often uses its conclusions without consciously worrying about whether the theorem is true.

is called the **standard deviation** of the random variable X. It is easily shown, in terms of definitions, that

$$\sigma^2 = E(X - \mu)^2 = E(X^2) - \mu^2;$$

that is, the variance can be written as the second moment about the origin minus the square of the mean.

It has already been shown that if $Z = g(X) = CX$, then $E(CX) = CE(X) = C\mu$, where C is any constant and μ is $E(X)$. The variance of the random variable $Z = g(X) = CX$ is also easily obtained. By definition, if X is a continuous random variable, the variance of Z is given by

$$E(Z - E(Z))^2 = E(CX - CE(X))^2 = \int_{-\infty}^{\infty} (Cy - C\mu)^2 f_X(y)\, dy$$
$$= C^2 \int_{-\infty}^{\infty} (y - \mu)^2 f_X(y)\, dy = C^2\sigma^2.$$

Thus, the variance of a constant times a random variable is just the square of the constant times the variance of the random variable. This is also true for discrete random variables. Finally, the variance of a constant is easily seen to be zero.

It has already been shown that if the demand for a product takes on the values 0, 1, 2, . . . , 99, each with probability $1/100$, then $E(X) = \mu = 49.5$. Similarly,

$$\sigma^2 = \sum_{k=0}^{99} (k - \mu)^2 P_X(k) = \sum_{k=0}^{99} k^2 P_X(k) - \mu^2$$
$$= \sum_{k=0}^{99} \frac{k^2}{100} - (49.5)^2 = 833.25.$$

Table 24.1 gives the means and variances of the random variables that are often useful in operations research. Note that for some random variables a single moment, the mean, provides a complete characterization of the distribution, e.g., the Poisson random variable. For some random variables the mean and variance provide a complete characterization of the distribution, e.g., the normal. In fact, if all the moments of a probability distribution are known, this is usually equivalent to specifying the entire distribution.

It was seen that the mean and variance may be sufficient to completely characterize a distribution, e.g., the normal. However, what can be said, in general, about a random variable whose mean μ and variance σ^2 are known, but nothing else about the form of the distribution is specified? This can be expressed in terms of **Chebyshev's inequality**, which states that for any positive number C,

$$P\{\mu - C\sigma \leq X \leq \mu + C\sigma\} > 1 - \frac{1}{C^2},$$

where X is any random variable having mean μ and variance σ^2. For example, if $C = 3$, if follows that $P\{\mu - 3\sigma \leq X \leq \mu + 3\sigma\} > 1 - 1/9 = 0.8889$. However, if X is known to have a normal distribution, then $P\{\mu - 3\sigma \leq X \leq \mu + 3\sigma\} = 0.9973$. Note that the Chebyshev inequality only gives a lower bound on the probability (usually a very conservative one), so there is no contradiction here.

24.9 BIVARIATE PROBABILITY DISTRIBUTION

Thus far the discussion has been concerned with the probability distribution of a single random variable, e.g., the demand for a product during the first month or the demand for a product during the second month. In an experiment that measures the demand during the first 2 months, it may well be important to look at the probability distribution of the vector random variable (X_1, X_2), the demand during the first month, and the demand during the second month, respectively,

Define the symbol

$$E_{b_1, b_2}^{X_1, X_2} = \{\omega | X_1(\omega) \leq b_1, X_2(\omega) \leq b_2\},$$

or equivalently,

$$E_{b_1, b_2}^{X_1, X_2} = \{X_1 \leq b_1, X_2 \leq b_2\},$$

as the set of outcomes ω in the sample space forming the event $E_{b_1, b_2}^{X_1, X_2}$, such that the random variable X_1 taken on values less than or equal to b_1, and X_2 takes on values less than or equal to b_2. Then $P\{E_{b_1, b_2}^{X_1, X_2}\}$ denotes the probability of this event. In the above example of the demand for a product during the first 2 months, suppose that the sample space Ω consists of the set of all possible points ω, where ω represents a pair of nonnegative integer values (x_1,x_2). Assume that x_1 and x_2 are bounded by 99. Thus, there are $(100)^2\omega$ points in Ω. Suppose further that each point ω has associated with it a probability equal to $1/(100)^2$, except for the points $\omega = (0,0)$ and $\omega = (99,99)$. The probability associated with the event $\{0,0\}$ will be $1.5/(100)^2$, that is, $P\{0,0\} = 1.5/(100)^2$, and the probability associated with the event $\{99,99\}$ will be $0.5/(100)^2$; that is, $P\{99,99\} = 0.5/(100)^2$. Thus, if there is interest in the "bivariate" random variable (X_1, X_2), the demand during the first and second months, respectively, then the event

$$\{X_1 \leq 1, X_2 \leq 3\}$$

is the set

$$E_{1,3}^{X_1, X_2} = \{(0,0), (0,1), (0,2), (0,3), (1,0), (1,1), (1,2), (1,3)\}.$$

Furthermore,

$$\begin{aligned} P\{E_{1,3}^{X_1, X_2}\} &= \frac{1.5}{(100)^2} + \frac{1}{(100)^2} + \frac{1}{(100)^2} + \frac{1}{(100)^2} + \frac{1}{(100)^2} + \frac{1}{(100)^2} + \frac{1}{(100)^2} \\ &\quad + \frac{1}{(100)^2} \\ &= \frac{8.5}{(100)^2}, \end{aligned}$$

so that

$$P\{X_1 \leq 1, X_2 \leq 3\} = P\{E_{1,3}^{X_1, X_2}\} = \frac{8.5}{(100)^2}.$$

A similar calculation can be made for any value of b_1 and b_2.

For any given bivariate random variable (X_1, X_2), $P\{X_1 \leq b_1, X_2 \leq b_2\}$ is denoted by $F_{X_1X_2}(b_1,b_2)$ and is called the **joint cumulative distribution function** (CDF) of the bivariate random variable (X_1, X_2) and is defined for all real values of b_1 and b_2. Where there is no ambiguity the joint CDF may be denoted by $F(b_1, b_2)$. Thus, attached to every bivariate random variable is a joint CDF. This is not an arbitrary function but is induced by the probabilities associated with events defined over the sample space Ω such that $\{\omega | X_1(\omega) \leq b_1, X_2(\omega) \leq b_2\}$.

The joint CDF of a random variable is a numerically valued function, defined for all b_1, b_2 such that $-\infty \leq b_1, b_2 \leq \infty$, having the following properties:

1. $F_{X_1X_2}(b_1,\infty) = P\{X_1 \leq b_1, X_2 \leq \infty\} = P\{X_1 \leq b_1\} = F_{X_1}(b_1)$, where $F_{X_1}(b_1)$ is just the CDF of the univariate random variable X_1.
2. $F_{X_1X_2}(\infty,b_2) = P\{X_1 \leq \infty, X_2 \leq b_2\} = P\{X_2 \leq b_2\} = F_{X_2}(b_2)$, where $F_{X_2}(b_2)$ is just the CDF of the univariate random variable X_2.

■ **TABLE 24.1** Table of common distributions

Distribution of random variable X	Form	Para-meters	Expected value	Variance	Range of random variable
Binomial	$P_X(k) = \frac{n!}{k!(n-k)!} p^k(1-p)^{n-k}$	n, p	np	$np(1-p)$	$0, 1, 2, \ldots, n$
Poisson	$P_X(k) = \frac{\lambda^k e^{-\lambda}}{k!}$	λ	λ	λ	$0, 1, 2, \ldots.$
Geometric	$P_X(k) = p(1-p)^{k-1}$	p	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$1, 2, \ldots.$
Exponential	$f_X(y) = \frac{1}{\theta} e^{-y/\theta}$	θ	θ	θ^2	$(0,\infty)$
Gamma	$f_X(y) = \frac{1}{\Gamma(\alpha)\beta^\alpha} y^{(\alpha-1)}e^{-y/\beta}$	α, β	$\alpha\beta$	$\alpha\beta^2$	$(0,\infty)$
Beta	$f_X(y) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{(\alpha-1)}(1-y)^{(\beta-1)}$	α, β	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	$(0,1)$
Normal	$f_X(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(y-\mu)^2/2\sigma^2}$	μ, σ	μ	σ^2	$(-\infty,\infty)$
Students t	$f_X(y) = \frac{1}{\sqrt{2\pi\nu}} \frac{\Gamma([\nu+1]/2)}{\Gamma(\nu/2)}(1+y^2/\nu)^{-(\nu+1)/2}$	ν	$0(\text{for } \nu > 1)$	$\nu/(\nu-2)(\text{for } \nu > 2)$	$(-\infty,\infty)$
Chi square	$f_X(y) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} y^{(\nu-2)/2}e^{-y/2}$	ν	ν	2ν	$(0,\infty)$
F	$f_X(y) = \frac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)\nu_1^{\nu_1/2}\nu_2^{\nu_2/2}}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} \frac{(y)^{(\nu_1/2)-1}}{(\nu_2+\nu_1 y)^{(\nu_1+\nu_2)/2}}$	ν_1,ν_2	$\frac{\nu_2}{\nu_2-2}$ for $\nu_2 > 2.$	$\frac{\nu_2^2(2\nu_2+2\nu_1-4)}{\nu_1(\nu_2-2)^2(\nu_2-4)}$ for $\nu_2 > 4$	$(0,\infty)$

3. $F_{X_1X_2}(b_1, -\infty) = P\{X_1 \le b_1, X_2 \le -\infty\} = 0$,
$F_{X_1X_2}(-\infty, b_2) = P\{X_1 \le -\infty, X_2 \le b_2\} = 0$.

4. $F_{X_1X_2}(b_1 + \Delta_1, b_2 + \Delta_2) - F_{X_1X_2}(b_1 + \Delta_1, b_2) - F_{X_1X_2}(b_1, b_2 + \Delta_2) + F_{X_1X_2}(b_1, b_2) \ge 0$, for every $\Delta_1, \Delta_2 \ge 0$, and b_1, b_2.

Using the definition of the event $E_{b_1, b_2}^{X_1, X_2}$, events of the form

$$\{a_1 < X_1 \le b_1, a_2 < X_2 \le b_2\}$$

can be described as the set of outcomes ω in the sample space such that the bivariate random variable (X_1, X_2) takes on values such that X_1 is greater than a_1 but does not exceed b_1 *and* X_2 is greater than a_2 but does not exceed b_2. $P\{a_1 < X_1 \le b_1, a_2 < X_2 \le b_2\}$ can easily be seen to be

$$F_{X_1X_2}(b_1, b_2) - F_{X_1X_2}(b_1, a_2) - F_{X_1X_2}(a_1, b_2) + F_{X_1X_2}(a_1, a_2).$$

It was noted that single random variables are generally characterized as discrete or continuous random variables. A bivariate random variable can be characterized in a similar manner. A bivariate random variable (X_1, X_2) is called a discrete bivariate random variable if both X_1 and X_2 are discrete random variables. Similarly, a bivariate random variable (X_1, X_2) is called a continuous bivariate random variable if both X_1 and X_2 are continuous random variables. Of course, bivariate random variables that are neither discrete nor continuous can exist, but these will not be important in this book.

The joint CDF for a discrete random variable $F_{X_1X_2}(b_1, b_2)$ is given by

$$\begin{aligned} F_{X_1X_2}(b_1, b_2) &= P\{\omega \mid X_1(\omega) \le b_1, X_2(\omega) \le b_2\} \\ &= \sum_{\text{all } k \le b_1} \sum_{\text{all } l \le b_2} P\{\omega \mid X_1(\omega) = k, X_2(\omega) = l\} \\ &= \sum_{\textit{all } k \le b_1} \sum_{\text{all } l \le b_2} P_{X_1X_2}(k, l), \end{aligned}$$

where $\{\omega \mid X_1(\omega) = k, X_2(\omega) = l)$ is the set of outcomes ω in the sample space such that the random variable X_1 taken on the value k and the variable X_2 takes on the value l; and $P\{\omega \mid X_1(\omega) = k, X_2(\omega) = l\} = P_{X_1X_2}(k, l)$ denotes the probability of this event. The $\mathrm{P}_{X_1X_2}(k, l)$ are called the joint probability distribution of the discrete bivariate random variable (X_1, X_2). Thus, in the example considered at the beginning of this section,

$$P_{X_1X_2}(k, 1) = 1/(100)^2 \text{ for all } k, l \text{ that are integers between 0 and 99,}$$

except for $P_{X_1X_2}(0, 0) = 1.5/(100)^2$ and $P_{X_1X_2}(99,99) = 0.5/(100)^2$.

For a continuous random variable, the joint CDF $F_{X_1X_2}(b_1, b_2)$ can usually be written as

$$F_{X_1X_2}(b_1,b_2) = P\{\omega \mid X_1(\omega) \le b_1, X_2(\omega) \le b_2\} = \int_{-\infty}^{b_1} \int_{-\infty}^{b_2} f_{X_1X_2}(s, t)\, ds\, dt,$$

where $f_{X_1X_2}(s, t)$ is known as the joint density function of the bivariate random variable (X_1, X_2). A knowledge of the joint density function enables one to calculate all sorts of probabilities, for example.

$$P\{a_1 < X_1 \le b_1, a_2 < X_2 \le b_2\} = \int_{a_1}^{b_1} \int_{a_2}^{b_2} f_{X_1X_2}(s, t)\, ds\, dt.$$

Finally, if the density function is known, it is said that the probability distribution of the random variable is determined.

The joint density function can be viewed as a surface in three dimensions, where the volume under this surface over regions in the s, t plane correspond to probabilities. Naturally, the density function can be obtained from the CDF by using the relation

$$\frac{\partial^2 F_{X_1X_2}(s,t)}{\partial s\,\partial t} = \frac{\partial^2}{\partial s\,\partial t}\int_{-\infty}^{s}\int_{-\infty}^{t} f_{X_1X_2}(u,v)\,du\,dv = f_{X_1X_2}(s,t).$$

In defining the joint CDF for a bivariate random variable, it was implied that $f_{X_1X_2}(s, t)$ was defined over the entire plane because

$$F_{X_1X_2}(b_1, b_2) = \int_{-\infty}^{b_1}\int_{-\infty}^{b_2} f_{X_1X_2}(s,t)\,ds\,dt$$

(which is analogous to what was done for a univariate random variable). This causes no difficulty, even for bivariate random variables having one or more components that cannot take on negative values or are restricted to other regions. In this case, $f_{X_1X_2}(s, t)$ can be defined to be zero over the inadmissible part of the plane. In fact, the only requirements for a function to be a bivariate density function are that

1. $f_{X_1X_2}(s, t)$ be nonnegative, and
2. $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X_1X_2}(s,t)\,ds\,dt = 1.$

24.10 MARGINAL AND CONDITIONAL PROBABILITY DISTRIBUTIONS

In Sec. 24.9 the discussion was concerned with the joint probability distribution of a bivariate random variable (X_1,X_2). However, there may also be interest in the probability distribution of the random variables X_1 and X_2 considered separately. It was shown that if $F_{X_1X_2}(b_1, b_2)$ represents the joint CDF of (X_1,X_2), then $F_{X_1}(b_1) = F_{X_1X_2}(b_1, \infty) = P\{X_1 \le b_1, X_2 \le \infty\} = P\{X_1 \le b_1\}$ is the CDF for the univariate random variable X_1, and $F_{X_2}(b_2) = F_{X_1X_2}(\infty, b_2) = P\{X_1 \le \infty, X_2 \le b_2\} = P\{X_2 \le b_2\}$ is the CDF for the univariate random variable X_2.

If the bivariate random variable (X_1, X_2) is discrete, it was noted that the

$$P_{X_1X_2}(k, l) = P\{X_1 = k, X_2 = l\}$$

describe its joint probability distribution. The probability distribution of X_1 individually, $P_{X_1}(k)$, now called the **marginal probability distribution** of the discrete random variable X_1, can be obtained from the $P_{X_1X_2}(k, l)$. In particular,

$$F_{X_1}(b_1) = F_{X_1X_2}(b_1,\infty) = \sum_{\text{all } k \le b_1}\sum_{\text{all } l} P_{X_1X_2}(k,l) = \sum_{\text{all } k \le b_1} P_{X_1}(k),$$

so that

$$P_{X_1}(k) = P\{X_1 = k\} = \sum_{\text{all } l} P_{X_1X_2}(k,l).$$

Similarly, the marginal probability distribution of the discrete random variable X_2 is given by

$$P_{X_2}(l) = P\{X_2 = l\} = \sum_{\text{all } k} P_{X_1X_2}(k,l).$$

Consider the experiment described in Sec. 24.1 which measures the demand for a product during the first 2 months, but where the probabilities are those given at the beginning of Sec. 24.9. The marginal distribution of X_1 is given by

$$\begin{aligned} PX_1(0) &= \sum_{\text{all } l} P_{X_1X_2}(0,l) \\ &= P_{X_1X_2}(0,0) + P_{X_1X_2}(0,1) + \cdots + P_{X_1X_2}(0,99) \\ &= \frac{1.5}{(100)^2} + \frac{1}{(100)^2} + \cdots + \frac{1}{(100)^2} = \frac{100.5}{(100)^2}, \end{aligned}$$

$$P_{X_1}(1) = P_{X_1}(2) = \cdots = P_{X_1}(98) = \sum_{\text{all } l} P_{X_1X_2}(k, l)$$
$$= \frac{100}{(100)^2}, \text{ for } k = 1, 2, \ldots, 98.$$
$$P_{X_1}(99) = \sum_{\text{all } l} P_{X_1X_2}(99, l)$$
$$= P_{X_1X_2}(99,0) + P_{X_1X_2}(99,1) + \cdots + P_{X_1X_2}(99,99)$$
$$= \frac{1}{(100)^2} + \frac{1}{(100)^2} + \cdots + \frac{0.5}{(100)^2} = \frac{99.5}{(100)^2}.$$

Note that this is indeed a probability distribution in that

$$P_{X_1}(0) + P_{X_1}(1) + \cdots + P_{X_1}(99) = \frac{100.5}{(100)^2} + \frac{100}{(100)^2} + \cdots + \frac{99.5}{(100)^2} = 1.$$

Similarly, the marginal distribution of X_2 is given by

$$P_{X_2}(0) = \sum_{\text{all } k} P_{X_1X_2}(k, 0)$$
$$= P_{X_1X_2}(0,0) + P_{X_1X_2}(1,0) + \cdots + P_{X_1X_2}(99,0)$$
$$= \frac{1.5}{(100)^2} + \frac{1}{(100)^2} + \cdots + \frac{1}{(100)^2} = \frac{100.5}{(100)^2},$$
$$P_{X_2}(1) = P_{X_2}(2) = \cdots = P_{X_2}(98) = \sum_{\text{all } k} P_{X_1X_2}(k, l) = \frac{100}{(100)^2}, \; l = 1, 2, \ldots, 98,$$
$$P_{X_2}(99) = \sum_{\text{all } k} P_{X_1X_2}(k, 99)$$
$$= P_{X_1X_2}(0,99) + P_{X_1X_2}(1,99) + \cdots + P_{X_1X_2}(99,99)$$
$$= \frac{1}{(100)^2} + \frac{1}{(100)^2} + \cdots + \frac{0.5}{(100)^2} = \frac{99.5}{(100)^2}.$$

If the bivariate random variable (X_1, X_2) is continuous, then $f_{X_1X_2}(s, t)$ represents the **joint density.** The density function of X_1 individually, $f_{X_1}(s)$, now called the **marginal density function** of the continuous random variable X_1, can be obtained from the $f_{X_1X_2}(s, t)$. In particular,

$$F_{X_1}(b_1) = F_{X_1X_2}(b_1, \infty) = \int_{-\infty}^{b_1} \int_{-\infty}^{\infty} f_{X_1X_2}(s, t)\, dt\, ds = \int_{-\infty}^{b_1} f_{X_1}(s)\, ds,$$

so that

$$f_{X_1}(s) = \int_{-\infty}^{\infty} f_{X_1X_2}(s, t)\, dt.$$

Similarly, the marginal density function of the continuous random variable X_2 is given by

$$f_{X_2}(t) = \int_{-\infty}^{\infty} f_{X_1X_2}(s, t)\, ds.$$

As indicated in Section 24.4, experiments are often performed where some results are obtained early in time and further results later in time. For example, in the previously described experiment that measures the demand for a product during the first two months, the demand for the product during the first month is observed at the end of the first month. This information can be utilized in making probability statements about the demand during the second month.

In particular, if the bivariate random variable (X_1, X_2) is discrete, the conditional probability distribution of X_2, given X_1, can be defined as

$$P_{X_2|X_1=k}(l) = P\{X_2 = l|X_1 = k\} = \frac{P_{X_1X_2}(k, l)}{P_{X_1}(k)}, \text{ if } P_{X_1}(k) > 0,$$

and the conditional probability distribution of X_1, given X_2, as

$$P_{X_1|X_2=l}(k) = P\{X_1 = k|X_2 = l\} = \frac{P_{X_1X_2}(k, l)}{P_{X_2}(l)}, \text{ if } P_{X_2}(l) > 0.$$

Note that for a given $X_2 = l$, $P_{X_1|X_2=l}(k)$ satisfies all the conditions for a probability distribution for a discrete random variable. $P_{X_1|X_2=l}(k)$ is nonnegative, and furthermore,

$$\sum_{\text{all } k} P_{X_1|X_2=l}(k) = \sum_{\text{all } k} \frac{P_{X_1X_2}(k, l)}{P_{X_2}(l)} = \frac{P_{X_2}(l)}{P_{X_2}(l)} = 1.$$

Again, returning to the demand for a product during the first 2 months, if it were known that there was no demand during the first month, then

$$P_{X_2|X_1=0}(l) = P\{X_2 = l|X_1 = 0\} = \frac{P_{X_1X_2}(0, l)}{P_{X_1}(0)} = \frac{P_{X_1X_2}(0, l)}{100.5/(100)^2}.$$

Hence,

$$P_{X_2|X_1=0}(0) = \frac{P_{X_1X_2}(0,0)}{(100.5)/(100)^2} = \frac{1.5}{100.5},$$

and

$$P_{X_2|X_1=0}(l) = \frac{1}{100.5} \qquad l = 1, 2, \ldots, 99.$$

If the bivariate random variable (X_1, X_2) is continuous with joint density function $f_{X_1X_2}(s, t)$, and the marginal density function of X_1 is given by $f_{X_1}(s)$, then the **conditional density function** of X_2, given $X_1 = s$, is defined as

$$f_{X_2|X_1=s}(t) = \frac{f_{X_1X_2}(s, t)}{f_{X_1}(s)}, \quad \text{if } f_{X_1}(s) > 0.$$

Similarly, if the marginal density function of X_2 is given by $f_{X_2}(t)$, then the conditional density function of X_1, given $X_2 = t$, is defined as

$$f_{X_1|X_2=t}(s) = \frac{f_{X_1X_2}(s, t)}{f_{X_2}(t)}, \quad \text{if } f_{X_2}(t) > 0.$$

Note that, given $X_1 = s$ and $X_2 = t$, the conditional density functions, $f_{X_2|X_1=s}(t)$ and $f_{X_1|X_2=t}(s)$, respectively, satisfy all the conditions for a density function. They are nonnegative, and furthermore,

$$\int_{-\infty}^{\infty} f_{X_2|X_1=s}(t)\, dt = \int_{-\infty}^{\infty} \frac{f_{X_1X_2}(s, t)\, dt}{f_{X_1}(s)}$$

$$= \frac{1}{f_{X_1}(s)} \int_{-\infty}^{\infty} f_{X_1X_2}(s, t)\, dt = \frac{f_{X_1}(s)}{f_{X_1}(s)} = 1,$$

and

$$\int_{-\infty}^{\infty} f_{X_1|X_2=t}(s)\, ds = \int_{-\infty}^{\infty} \frac{f_{X_1X_2}(s, t)\, ds}{f_{X_2}(t)}$$

$$= \frac{1}{f_{X_2}(t)} \int_{-\infty}^{\infty} f_{X_1X_2}(s, t)\, ds = \frac{f_{X_2}(t)}{f_{X_2}(t)} = 1.$$

As an example of the use of these concepts for a continuous bivariate random variable, consider an experiment that measures the time of the first arrivals at a store on each of two successive days. Suppose that the joint density function for the random variable (X_1, X_2), which represents the arrival time on the first and second days, respectively, is given by

$$f_{X_1X_2}(s, t) = \begin{cases} \dfrac{1}{\theta^2}\, e^{-(s+t)/\theta}, & \text{for } s, t \geq 0 \\ 0, & \text{otherwise.} \end{cases}$$

The marginal density function of X_1 is given by

$$f_{X_1}(s) = \begin{cases} \displaystyle\int_0^\infty \frac{1}{\theta^2}\, e^{-(s+t)/\theta} dt = \frac{1}{\theta} e^{-s/\theta}, & \text{for } s \geq 0 \\ 0, & \text{otherwise.} \end{cases}$$

and the marginal density function of X_2 is given by

$$f_{X_2}(t) = \begin{cases} \displaystyle\int_0^\infty \frac{1}{\theta^2}\, e^{-(s+t)/\theta}\, ds = \frac{1}{\theta} e^{-t/\theta}, & \text{for } t \geq 0 \\ 0, & \text{otherwise.} \end{cases}$$

If it is announced that the arrival time of the first customer on the first day occurred at time s, the conditional density of X_2, given $X_1 = s$, is given by

$$f_{X_2|X_1=s}(t) = \frac{f_{X_1X_2}(s, t)}{f_{X_1}(s)} = \frac{(1/\theta^2)e^{-(s+t)/\theta}}{(1/\theta)e^{-s/\theta}} = \frac{1}{\theta}\, e^{-t/\theta}.$$

It is interesting to note at this point that the conditional density of X_2, given $X_1 = s$, is independent of s and, furthermore, is the same as the marginal density of X_2.

24.11 EXPECTATIONS FOR BIVARIATE DISTRIBUTIONS

Section 24.7 defined the expectation of a function of a univariate random variable. The expectation of a function of a bivariate random variable (X_1, X_2) may be defined in a similar manner. Let $g(X_1, X_2)$ be a function of the bivariate random variable (X_1, X_2). Let

$$P_{X_1X_2}(k, l) = P\{X_1 = k, X_2 = l\}$$

denote the joint probability distribution if (X_1, X_2) is a discrete random variable, and let $f_{X_1X_2}(s, t)$ denote the joint density function if (X_1, X_2) is a continuous random variable. The expectation of $g(X_1, X_2)$ is now defined as

$$E[g(X_1, X_2)] = \begin{cases} \displaystyle\sum_{\text{all } k,l} g(k, l) P_{X_1X_2}(k, l), & \text{if } X_1, X_2 \text{ is a discrete random variable} \\ \displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(s, t) f_{X_1X_2}(s, t)\, ds\, dt, & \text{if } X_1, X_2 \text{ is a continuous random variable.} \end{cases}$$

An alternate definition can be obtained by recognizing that $Z = g(X_1, X_2)$ is itself a univariate random variable and hence has a density function if Z is continuous and a probability distribution if Z is discrete. The expectation of Z for these cases has already been defined in Sec. 24.7. Of particular interest here is the extension of the theorem of the unconscious statistician, which states that if (X_1, X_2) is a continuous random variable and if Z has a density function $h_Z(y)$, then

$$E(Z) = \int_{-\infty}^{\infty} yh_z(y)\,dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(s, t)f_{X_1X_2}(s, t)\,ds\,dt.$$

Thus, the expectation of Z can be found by using its definition in terms of the density of the univariate random variable Z or, alternatively, by use of its definition as the expectation of a function of the bivariate random variable (X_1, X_2) with respect to its joint density function. The identical theorem is true for a discrete bivariate random variable, and, of course, both results are easily extended to n-variate random variables.

There are several important functions g that should be considered. All the results will be stated for continuous random variables, but equivalent results also hold for discrete random variables.

If $g(X_1, X_2) = X_1$, it is easily seen that

$$E(X_1) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} sf_{X_1X_2}(s, t)\,ds\,dt = \int_{-\infty}^{\infty} sf_{X_1}(s)\,ds.$$

Note that this is just the expectation of the univariate random variable X_1 with respect to its marginal density.

In a similar manner, if $g(X_1, X_2) = [X_1 - E(X_1)]^2$, then

$$\begin{aligned} E[X_1 - E(X_1)]^2 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [s - E(X_1)]^2 f_{X_1X_2}(s, t)\,ds\,dt \\ &= \int_{-\infty}^{\infty} [s - E(X_1)]^2 f_{X_1}(s)\,ds, \end{aligned}$$

which is just the variance of the univariate random variable X_1 with respect to its marginal density.

If $g(X_1, X_2) = [X_1 - E(X_1)]\,[X_2 - E(X_2)]$, then $E[g(X_1, X_2)]$ is called the **covariance** of the random variable (X_1, X_2); that is,

$$E[X_1 - E(X_1)][X_2 - E(X_2)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [s - E(X_1)][t - E(X_2)]f_{X_1X_2}(s, t)\,ds\,dt.$$

An easy computational formula is provided by the identity

$$E[X_1 - E(X_1)][X_2 - E(X_2)] = E(X_1X_2) - E(X_1)E(X_2).$$

The **correlation coefficient** between X_1 and X_2 is defined to be

$$\rho = \frac{E[X_1 - E(X_1)][X_2 - E(X_2)]}{\sqrt{E[X_1 - E(X_1)]^2 E[X_2 - E(X_2)]^2}}.$$

It is easily shown that $-1 \le \rho \le +1$.

The final results pertain to a linear combination of random variables. Let $g(X_1, X_2) = C_1X_1 + C_2X_2$, where C_1 and C_2 are constants. Then

$$\begin{aligned} E[g(X_1, X_2)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (C_1 s + C_2\,t)\,f_{X_1X_2}(s, t)\,ds\,dt, \\ &= C_1\int_{-\infty}^{\infty} sf_{X_1}(s)\,ds + C_2\int_{-\infty}^{\infty} tf_{X_2}(t)\,dt, \\ &= C_1E(X_1) + C_2E(X_2). \end{aligned}$$

Thus, the expectation of a linear combination of univariate random variables is just

$$E[C_1X_1 + C_2X_2 + \cdots + C_nX_n] = C_1E(X_1) + C_2E(X_2) + \cdots + C_nE(X_n).$$

If

$$g(X_1, X_2) = [C_1X_1 + C_2X_2 - \{C_1E(X_1) + C_2E(X_2)\}]^2,$$

then

$$\begin{aligned} E[g(X_1, X_2)] &= \text{variance } (C_1X_1 + C_2 X_2) \\ &= C_1^2E[X_1 - E(X_1)]^2 + C_2^2E[X_2 - E(X_2)]^2 \\ &\quad + 2C_1C_2E[X_1 - E(X_1)][X_2 - E(X_2)] \\ &= C_1^2 \text{ variance } (X_1) + C_2^2 \text{ variance } (X_2) \\ &\quad + 2C_1C_2 \text{ covariance } (X_1X_2). \end{aligned}$$

For n univariate random variables, the variance of a linear combination $C_1X_1 + C_2 X_2 + \cdots + C_nX_n$ is given by

$$\sum_{i=1}^{n} C_i^2 \text{ variance } (X_i) + 2\sum_{j=2}^{n}\sum_{i=1}^{j-1} C_iC_j \text{ covariance } (X_iX_j).$$

24.12 INDEPENDENT RANDOM VARIABLES AND RANDOM SAMPLES

The concept of independent events has already been defined; that is, E_1 and E_2 are independent events if, and only if,

$$P\{E_1 \cap E_2\} = P\{E_1\}P\{E_2\}.$$

From this definition the very important concept of *independent random variables* can be introduced. For a bivariate random variable (X_1,X_2) and constants b_1 and b_2, denote by E_1 the event containing those ω such that $X_1(\omega) \leq b_1$, $X_2(\omega)$ is anything; that is,

$$E_1 = \{\omega \,|\, X_1(\omega) \leq b_1, X_2(\omega) \leq \infty\}.$$

Similarly, denote by E_2 the event containing those ω such that $X_1(\omega)$ is anything and $X_2(\omega) \leq b_2$; that is,

$$E_2 = \{\omega \,|\, X_1(\omega) \leq \infty, X_2(\omega) \leq b_2\}.$$

Furthermore, the event $E_1 \cap E_2$ is given by

$$E_1 \cap E_2 = \{\omega \,|\, X_1(\omega) \leq b_1, X_2(\omega) \leq b_2\}.$$

The random variables X_1 and X_2 are said to be independent if events of the form given by E_1 and E_2 are independent events for all b_1 and b_2. Using the definition of independent events, then, the random variables X_1 and X_2 are called **independent random variables** if

$$P\{X_1 \leq b_1, X_2 \leq b_2\} = P\{X_1 \leq b_1\}P\{X_2 \leq b_2\}$$

for all b_1 and b_2. Therefore, X_1 and X_2 are independent if

$$\begin{aligned} F_{X_1X_2}(b_1, b_2) &= P\{X_1 \leq b_1, X_2 \leq b_2\} = P\{X_1 \leq b_1\}P\{X_2 \leq b_2\} \\ &= F_{X_1}(b_1)F_{X_2}(b_2). \end{aligned}$$

Thus, the independence of the random variables X_1 and X_2 implies that the joint CDF factors into the product of the CDF's of the individual random variables. Furthermore, it is easily shown that if (X_1,X_2) is a discrete bivariate random variable, then X_1 and X_2 are independent random variables if, and only if, $P_{X_1X_2}(k, l) = P_{X_1}(k)P_{X_2}(l)$; in other words, $P\{X_1 = k, X_2 = l\} = P\{X_1 = k\}P\{X_2 = l\}$, for all k and l. Similarly, if (X_1, X_2) is a continuous bivariate random variable, then X_1 and X_2 are independent random variables if, and only if,

$$f_{X_1X_2}(s, t) = f_{X_1}(s)f_{X_2}(t),$$

for all s ant t. Thus, if X_1, X_2 are to be independent random variables, the joint density (or probability) function must factor into the product of the marginal density functions of the random variables. Using this result, it is easily seen that if X_1, X_2 are independent random variables, then the covariance of X_1, X_2 must be zero. Hence, the results on the variance of linear combinations of random variables given in Sec. 24.11 can be simplified when the random variables are independent; that is,

$$\text{Variance}\left(\sum_{i=1}^{n} C_i X_i\right) = \sum_{i=1}^{n} C_i^2 \text{ variance } (X_i)$$

when the X_i are independent.

Another interesting property of independent random variables can be deduced from the factorization property. If (X_1, X_2) is a discrete bivariate random variable, then X_1 and X_2 are independent if, and only if,

$$P_{X_1|X_2=l}(k) = P_{X_1}(k), \text{ for all } k \text{ and } l.$$

Similarly, if (X_1, X_2) is a continuous bivariate random variable, then X_1 and X_2 are independent if, and only if,

$$f_{X_1|X_2=t}(s) = f_{X_1}(s), \text{ for all } s \text{ and } t.$$

In other words, if X_1 and X_2 are independent, a knowledge of the outcome of one, say, X_2, gives no information about the probability distribution of the other, say, X_1. It was noted in the example in Sec. 24.10 on the time of first arrivals that the conditional density of the arrival time of the first customer on the second day, given that the first customer on the first day arrived at time s, was equal to the marginal density of the arrival time of the first customer on the second day. Hence, X_1 and X_2 were independent random variables. In the example of the demand for a product during two consecutive months with the probabilities given in Sec. 24.9, it was seen in Sec. 24.10 that

$$P_{X_2|X_1=0}(0) = \frac{1.5}{100.5} \neq P_{X_2}(0) = \frac{100.5}{(100)^2}.$$

Hence, the demands during each month were dependent (not independent) random variables.

The definition of independent random variables generally does not lend itself to determine whether or not random variables are independent in a probabilistic sense by looking at their outcomes. Instead, by analyzing the physical situation the experimenter usually is able to make a judgment about whether the random variables are independent by ascertaining if the outcome of one will affect the probability distribution of the other.

The definition of independent random variables is easily extended to three or more random variables. For example, if the joint CDF of the n-dimensional random variable $(X_1, X_2, \ldots, X_n)$ is given by $F_{X_1X_2 \cdots X_n}(b_1, b_2, \ldots, b_n)$ and $F_{X_1}(b_1)$, $F_{X_2}(b_2), \ldots,$ $F_{X_n}(b_n)$ represents the CDF's of the univariate random variables $X_1, X_2, \ldots, X_n$, respectively, then $X_1, X_2, \ldots, X_n$ are independent random variables if, and only if,

$$F_{X_1X_2 \cdots X_n}(b_1, b_2, \ldots, b_n) = F_{X_1}(b_1)F_{X_2}(b_2) \cdots F_{X_n}(b_n), \text{ for all } b_1, b_2, \ldots, b_n.$$

Having defined the concept of independent random variables, we can now introduce the term *random sample*. A **random sample** simply means a sequence of independent and identically distributed random variables. Thus, $X_1, X_2, \ldots, X_n$ constitute a random sample of size n if the X_i are independent and identically distributed random variables. For example, in Sec. 24.5 it was pointed out that if $X_1, X_2, \ldots, X_n$ are independent Bernoulli random variables, each with parameter p (that is, if the X's are a random sample), then the random variable

$$X = \sum_{i=1}^{n} X_i$$

has a binomial distribution with parameters n and p.

24.13 LAW OF LARGE NUMBERS

Section 24.7 pointed out that the mean of a random sample tends to converge to the expectation of the random variables as the sample size increases. In particular, suppose the random variable X, the demand for a product, may take on one of the possible values $k = 0, 1, 2, \ldots, 98, 99$, each with $P_X(k) = 1/100$ for all k. Then $E(X)$ is easily seen to be 49.5. If a random sample of size n is taken, i.e., the demands are observed for n days, with each day's demand being independent and identically distributed random variables, it was noted that the random variable $\overline{X}$ should take on a value close to 49.5 if n is large. This result can be stated precisely as the *law of large numbers*.

Law of Large Numbers

Let the random variables $X_1, X_2, \ldots, X_n$ be independent, identically distributed random variables (a random sample of size n), each having mean μ. Consider the random variable that is the sample mean $\overline{X}$:

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}.$$

Then for any constant $\varepsilon > 0$,

$$\lim_{n\to\infty} P\{|\overline{X} - \mu| > \varepsilon\} = 0.$$

The interpretation of the law of large numbers is that as the sample size increases, the probability is "close" to 1 that $\overline{X}$ is "close" to μ. Assuming that the variance of each X_i is $\sigma^2 < \infty$, this result is easily proved by using Chebyshev's inequality (stated in Sec. 24.8). Since each X_i has mean μ and variance σ^2, $\overline{X}$ also has mean μ, but its variance is σ^2/n. Hence, applying Chebyshev's inequality to the random variable $\overline{X}$, it is evident that

$$P\left\{\mu - \frac{C\sigma}{\sqrt{n}} \le \overline{X} \le \mu + \frac{C\sigma}{\sqrt{n}}\right\} > 1 - \frac{1}{C^2}.$$

This is equivalent to

$$P\left\{|\overline{X} - \mu| > \frac{C\sigma}{\sqrt{n}}\right\} < \frac{1}{C^2}.$$

Let $C\sigma/\sqrt{n} = \varepsilon$, so that $C = \varepsilon\sqrt{n}/\sigma$. Thus,

$$P\{|\overline{X} - \mu| > \varepsilon\} < \frac{\sigma^2}{\varepsilon^2 n},$$

so that

$$\lim_{n\to\infty} P\{|\overline{X} - \mu| > \varepsilon\} = 0,$$

as was to be proved.

24.14 CENTRAL LIMIT THEOREM

Section 24.6 pointed out that sums of independent normally distributed random variables are themselves normally distributed, and that even if the random variables are *not* normally distributed, the distribution of their sum still tends toward normality. This latter statement can be made precise by means of the *central limit theorem*.

Central Limit Theorem

Let the random variables $X_1, X_2, \ldots, X_n$ be independent with means $\mu_1, \mu_2, \ldots, \mu_n$, respectively, and variance $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$, respectively. Consider the random variable Z_n,

$$Z_n = \frac{\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mu_i}{\sqrt{\sum_{i=1}^{n} \sigma_i^2}} .$$

Then, under certain regularity conditions, Z_n is approximately normally distributed with zero mean and unit variance in the sense that

$$\lim_{n\to\infty} P\{Z_n \le b\} = \int_{-\infty}^{b} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy.$$

Note that if the X_i form a random sample, with each X_i having mean μ and variance σ^2, then $Z_n = (\overline{X} - \mu)\sqrt{n}/\sigma$.† Hence, sample means from random samples tend toward normality in the sense just described by the central limit theorem even if the X_i are not normally distributed.

It is difficult to give sample sizes beyond which the central limit theorem applies and approximate normality can be assumed for sample means. This, of course, does depend upon the form of the underlying distribution. From a practical point of view, moderate sample sizes, like 10, are often sufficient.

24.15 FUNCTIONS OF RANDOM VARIABLES

Section 24.7 introduced the theorem of the unconscious statistician and pointed out that if a function $Z = g(X)$ of a continuous random variable is considered, its expectation can be taken with respect to the density function $f_X(y)$ of X or the density function $h_Z(y)$ of Z. In discussing this choice, it was implied that the density function of Z was known. In general, then, given the cumulative distribution function $F_X(b)$ of a random variable X, there may be interest in obtaining the cumulative distribution function $H_Z(b)$ of a random variable $Z = g(X)$. Of course, it is always possible to go back to the sample space and determine $H_Z(b)$ directly from probabilities associated with the sample space. However, alternate methods for doing this are desirable.

If X is a discrete random variable, the values k that the random variable X takes on and the associated $P_X(k)$ are known. If $Z = g(X)$ is also discrete, denote by m the values that Z takes on. The probabilities $Q_Z(m) = P\{Z = m\}$ for all m are required. The general procedure is to enumerate for each m all the values of k such that

$$g(k) = m.$$

$Q_Z(m)$ is then determined as

$$Q_Z(m) = \sum_{\substack{\text{all } k \\ \text{such that} \\ g(k) = m}} P_X(k).$$

To illustrate, consider again the example involving the demand for a product in a single month. Let this random variable be noted by X, and let $k = 0, 1, \ldots, 99$ with $P_X(k) = 1/100$, for all k. Consider a new random variable Z that takes on the value of 0 if there is no

†Under these conditions the central limit theorem actually holds without assuming any other regularity conditions.

demand and 1 if there is *any* demand. This random variable maybe useful for determining whether any shipping is needed. The probabilities

$$Q_Z(0) \text{ and } Q_Z(1)$$

are required. If $m = 0$, the only value of k such that $g(k) = 0$ is $k = 0$. Hence,

$$Q_Z(0) = \sum_{\substack{\text{all } k \\ \text{such that} \\ g(k)=0}} P_X(k) = P_X(0) = \frac{1}{100}.$$

If $m = 1$, the values of k such that $g(k) = 1$ are $k = 1, 2, 3, \ldots, 98, 99$. Hence,

$$\begin{aligned} Q_Z(1) &= \sum_{\substack{\text{all } k \\ \text{such that} \\ g(k)=1}} P_X(k) \\ &= P_X(1) + P_X(2) + P_X(3) + \cdots + P_X(98) + P_X(99) = \frac{99}{100}. \end{aligned}$$

If X is a continuous random variable, then both the CDF $F_X(b)$ and the density function $f_X(y)$ may be assumed to be known. If $Z = g(X)$ is also a continuous random variable, either the CDF $H_Z(b)$ or the density function $h_Z(y)$ is sought. To find $H_Z(b)$, note that

$$H_Z(b) = P\{Z \le b\} = P\{g(X) \le b\} = P\{A\},$$

where A consists of all points such that $g(X) \le b$. Thus, $P\{A\}$ can be determined from the density function of CDF of the random variable X. For example, suppose that the CDF for the time of the first arrival in a store is given by

$$F_X(b) = \begin{cases} 1 - e^{-b/\theta}, & \text{for } b \ge 0 \\ 0, & \text{for } b < 0, \end{cases}$$

where $\theta > 0$. Suppose further that the random variable $Z = g(X) = X + 1$, which represents an hour after the first customer arrives, is of interest, and the CDF of Z, $H_Z(b)$, is desired. To find this CDF note that

$$\begin{aligned} H_Z(b) &= P\{Z \le b\} = P\{X + 1 \le b\} = P\{X \le b - 1\} \\ &= \begin{cases} 1 - e^{-(b-1)/\theta}, & \text{for } b \ge 1 \\ 0, & \text{for } b < 1. \end{cases} \end{aligned}$$

Furthermore, the density can be obtained by differentiating the CDF; that is,

$$h_Z(y) = \begin{cases} \dfrac{1}{\theta} e^{-(y-1)/\theta}, & \text{for } y \ge 1 \\ 0, & \text{for } y < 1. \end{cases}$$

Another technique can be used to find the density function directly if g(X) is monotone and differentiable; it can be shown that

$$h_Z(y) = f_X(s) \left| \frac{ds}{dy} \right|,$$

where s is expressed in terms of y. In the example, $Z = g(X) = X + 1$, so that y, the value the random variable Z takes on, can be expressed in terms of s, the value the random variable X takes on; that is, $y = g(s) = s + 1$. Thus,

$$s = y - 1, \quad f_X(s) = \frac{1}{\theta} e^{-s/\theta} = \frac{1}{\theta} e^{-(y-1)/\theta}, \quad \text{and } \frac{ds}{dy} = 1.$$

Hence,

$$h_Z(y) = \frac{1}{\theta}\, e^{-(y-1)/\theta}\ |1| = \frac{1}{\theta}\, e^{-(y-1)/\theta},$$

which is the result previously obtained.

All the discussion in this section concerned functions of a single random variable. If (X_1, X_2) is a bivariate random variable, there may be interest in the probability distribution of such functions as $X_1 + X_2$, X_1X_2, X_1/X_2, and so on. If (X_1, X_2) is discrete, the technique for single random variables is easily extended. A detailed discussion of the techniques available for continuous bivariate random variables is beyond the scope of this text; however, a few notions related to independent random variables will be discussed.

If (X_1, X_2) is a continuous bivariate random variable, and X_1 and X_2 are independent, then its joint density is given by

$$f_{X_1X_2}(s, t) = f_{X_1}(s)f_{X_2}(t).$$

Consider the function

$$Z = g(X_1, X_2) = X_1 + X_2.$$

The CDF for Z can be expressed as $H_Z(b) = P\{Z \le b\} = P\{X_1 + X_2 \le b\}$. This can be evaluated by integrating the bivariate density over the region such that $s + t \le b$; that is

$$\begin{aligned} H_Z(b) &= \iint_{s+t \le b} f_{X_1}(s)f_{X_2}(t)\, ds\, dt \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{b-t} f_{X_1}(s)f_{X_2}(t)\, ds\, dt. \end{aligned}$$

Differentiating with respect to b yields the density function

$$h_Z(y) = \int_{-\infty}^{\infty} f_{X_2}(t)f_{X_1}(y - t)\, dt.$$

This can be written alternately as

$$h_Z(y) = \int_{-\infty}^{\infty} f_{X_1}(s)f_{X_2}(y - s)\, ds.$$

Note that the integrand may be zero over part of the range of the variable, as shown in the following example.

Suppose that the times of the first arrival on two successive days, X_1 and X_2, are independent, identically distributed random variables having density

$$f_{X_1}(s) = \begin{cases} \dfrac{1}{\theta}e^{-s/\theta}, & \text{for } s \ge 0 \\ 0, & \text{otherwise.} \end{cases}$$

$$f_{X_2}(t) = \begin{cases} \dfrac{1}{\theta}e^{-t/\theta}, & \text{for } t \ge 0 \\ 0, & \text{otherwise.} \end{cases}$$

To find the density of $Z = X_1 + X_2$, note that

$$f_{X_1}(s) = \begin{cases} \dfrac{1}{\theta}e^{-s/\theta}, & \text{for } s \ge 0 \\ 0, & \text{for } s < 0, \end{cases}$$

and

$$f_{X_2}(y - s) = \begin{cases} \dfrac{1}{\theta}e^{-(y-s)/\theta}, & \text{if } y-s \ge 0 \text{ so that } s \le y \\ 0, & \text{if } y-s < 0 \text{ so that } s > y. \end{cases}$$

Hence,

$$f_{X_1}(s)\, f_{X_2}(y - s) = \begin{cases} \dfrac{1}{\theta}e^{-s/\theta}\,\dfrac{1}{\theta}e^{-(y-s)/\theta} = \dfrac{1}{\theta^2}e^{-y/\theta}, & \text{if } 0 \le s \le y \\ 0, & \text{otherwise.} \end{cases}$$

Hence,

$$h_Z(y) = \int_{-\infty}^{\infty} f_{X_1}(s) f_{X_2}(y - s)\, ds = \int_0^y \cdot \frac{1}{\theta^2}\, e^{-y/\theta}\, ds$$

$$= \frac{y}{\theta^2}\, e^{-y/\theta}.$$

Note that this is just a gamma distribution, with parameters $\alpha = 2$ and $\beta = \theta$. Hence, as indicated in Sec. 24.6, the sum of two independent, exponentially distributed random variables has a gamma distribution. This example illustrates how to find the density function for finite sums of independent random variables. Combining this result with those for univariate random variables leads to easily finding the density function of linear combinations of independent random variables.

A final result on the distribution of functions of random variables concerns functions of normally distributed random variables. The chi-square and the t and F distributions, introduced in Sec. 24.6, can be generated from functions of normally distributed random variables. These distributions are particularly useful in the study of statistics. In particular, let $X_1, X_2, \ldots, X_\nu$ be independent, normally distributed random variables having zero mean and unit variance. The random variable

$$\chi^2 = X_1^2 + X_2^2 + \cdots + X_\nu^2$$

can be shown to have a *chi-square distribution* with ν degrees of freedom. A random variable having a t distribution may be generated as follows. Let X be a normally distributed random variable having zero mean and unit variance and χ^2 be a chi-square random variable (independent of X) with ν degrees of freedom. The random variable

$$t = \frac{\sqrt{\nu}\mathrm{X}}{\sqrt{\chi^2}}$$

can be shown to have a *t distribution* with ν degrees of freedom. Finally, a random variable having an F distribution can be generated from a function of two independent chi-square random variables. Let χ_1^2 and χ_2^2 be independent chi-square random variables, with ν_1 and ν_2 degrees of freedom, respectively. The random variable

$$F = \frac{\chi_1^2/\nu_1}{\chi_2^2/\nu_1}$$

can be shown to have an *F distribution* with ν_1 and ν_2 degrees of freedom.

SELECTED REFERENCES

1. Asmussen, S.: *Applied Probability and Queues,* 2nd ed., Springer, New York, 2003.
2. Bhattacharya, R., and E.C. Waymire: *A Basic Course on Probability Theory,* Springer, New York, 2007.
3. Billingsley, P.: *Probability and Measure,* 4th ed., Wiley, Hoboken, NJ, 2012.
4. Durret, R.: *Probability Theory and Examples,* 4th ed., Cambridge University Press, Cambridge, UK, 2010.
5. Feller, W.: *An Introduction to Probability Theory and Its Applications,* vol. 1, 3d ed., Wiley, New York, 1968.
6. Feller, W.: *An Introduction to Probability Theory and Its Applications,* vol. 2, 2d ed., Wiley, New York, 1971.

7. Ross, S.: *A First Course in Probability,* 9th ed., Pearson, Upper Saddle River, NJ, 2014.
8. ____: *Introduction to Probability and Statistics for Engineers and Scientists,* 4th ed., Academic Press, Orlando, FL, 2010.
9. ____: *Introduction to Probability Models,* 10th ed., Academic Press, Orlando, FL, 2010.

PROBLEMS

24-1. A cube has its six sides colored red, white, blue, green, yellow, and violet. It is assumed that these six sides are equally likely to show when the cube is tossed. The cube is tossed once.

(a) Describe the sample space.
(b) Consider the random variable that assigns the number 0 to red and white, the number 1 to green and blue, and the number 2 to yellow and violet. What is the distribution of this random variable?
(c) Let $Y = (X + 1)^2$, where X is the random variable in part (b). Find $E(Y)$.

24-2. Suppose the sample space Ω consists of the four points

$$\omega_1, \omega_2, \omega_3, \omega_4,$$

and the associated probabilities over the events are given by

$$P\{\omega_1\} = \frac{1}{3}, P\{\omega_2\} = \frac{1}{5}, P\{\omega_3\} = \frac{3}{10}, P\{\omega_4\} = \frac{1}{6}.$$

Define the random variable X_1 by

$$X_1(\omega_1) = 1,$$
$$X_1(\omega_2) = 1,$$
$$X_1(\omega_3) = 4,$$
$$X_1(\omega_4) = 5,$$

and the random variable X_2 by

$$X_1(\omega_1) = 1,$$
$$X_2(\omega_2) = 1,$$
$$X_2(\omega_3) = 1,$$
$$X_2(\omega_4) = 5,$$

(a) Find the probability distribution of X_1, that is, $P_{X_1}(i)$.
(b) Find $E(X_1)$.
(c) Find the probability distribution of the random variable $X_1 + X_2$, that is, $P_{X_1+X_2}(i)$.
(d) Find $E(X_1 + X_2)$ and $E(X_2)$.
(e) Find $F_{X_1X_2}(b_1, b_2)$.
(f) Compute the correlation coefficient between X_1 and X_2.
(g) Compute $E[2X_1 - 3X_2]$.

24-3. During the course of a day a machine turns out two items, one in the morning and one in the afternoon. The quality of each item is measured as good (G), mediocre (M), or bad (B). The long-run fraction of good items the machine produces is $1/2$, the fraction of mediocre items is $1/3$, and the fraction of bad items is $1/6$.

(a) In a column, write the sample space for the experiment that consists of observing the day's production.
(b) Assume a good item returns a profit of \$2, a mediocre item a profit of \$1, and a bad item yields nothing. Let X be the random variable describing the total profit for the day. In a column adjacent to the column in part (a), write the value of this random variable corresponding to each point in the sample space.
(c) Assuming that the qualities of the morning and afternoon items are independent, in a third column associate with every point in the sample space a probability for that point.
(d) Write the set of all possible outcomes for the random variable X. Give the probability distribution function for the random variable.
(e) What is the expected value of the day's profit?

24-4. The random variable X has density function f given by

$$f_X(y) = \begin{cases} \theta, & \text{for } 0 \le y \le \theta \\ K, & \text{for } \theta < y \le 1 \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Determine K in terms of θ.
(b) Find $F_X(b)$, the CDF of X.
(c) Find $E(X)$.
(d) Suppose $\theta = \frac{1}{3}$. Is $P\left\{X - \frac{1}{3} < a\right\} = P\left\{-\left(X - \frac{1}{3}\right) < a\right\}$?

24-5. Let X be a discrete random variable, with probability distribution

$$P\{X = x_1\} = \frac{1}{4}$$

and

$$P\{X = x_2\} = \frac{3}{4}.$$

(a) Determine x_1 and x_2, such that

$$E(X) = 0 \text{ and variance } (X) = 10.$$

(b) Sketch the CDF of X.

24-6. The life X, in hours, of a certain kind of radio tube has a probability density function given by

$$f_X(y) = \begin{cases} \dfrac{100}{y^2}, & \text{for } y \ge 100 \\ 0, & \text{for } y < 100. \end{cases}$$

(a) What is the probability that a tube will survive 250 hours of operation?
(b) Find the expected value of the random variable.

24-7. The random variable X can take on only the values 0, ±1, ±2, and

$$P\{-1 < X < 2\} = 0.4, \qquad P\{X = 0\} = 0.3,$$
$$P\{|X| \leq 1\} = 0.6, \qquad P\{X \geq 2\} = P\{X = 1 \text{ or } -1\}.$$

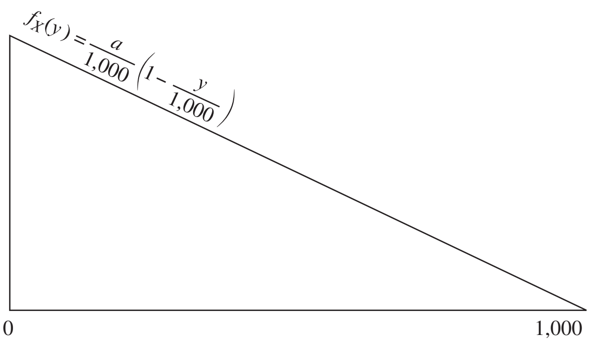
(a) Find the probability distribution of X.
(b) Graph the CDF of X.
(c) Compute $E(X)$.

24-8. Let X be a random variable with density

$$f_X(y) = \begin{cases} K(1 - y^2), & \text{for } -1 < y < 1 \\ 0, & \text{otherwise} \end{cases}$$

(a) What value of K will make $f_X(y)$ a true density?
(b) What is the CDF of X?
(c) Find $E(2X - 1)$.
(d) Find variance (X).
(e) Find the approximate value of $P\{\overline{X} > 0\}$, where $\overline{X}$ is the sample mean from a random sample of size $n = 100$ from the above distribution. (*Hint:* Note that n is "large.")

24-9. The distribution of X, the life of a transistor, in hours, is approximated by a triangular distribution as follows:



(a) What is the value of a?
(b) Find the expected value of the life of transistors.
(c) Find the CDF, $F_X(b)$, for this density. Note that this must be defined for all b between plus and minus infinity.
(d) If X represents the random variable, the life of a transistor, let $Z = 3X$ be a new random variable. Using the results of (c), find the CDF of Z.

24-10. The number of orders per week, X, for radios can be assumed to have a Poisson distribution with parameter $\lambda = 25$.
(a) Find $P\{X \geq 25\}$ and $P\{X = 20\}$.
(b) If the number of radios in the inventory is 35, what is the probability of a shortage occurring in a week?

24-11. Consider the following game. Player A flips a fair coin until a head appears. She pays player B 2^n dollars, where n is the number of tosses required until a head appears. For example, if a head appears on the first trial, player A pays player B \$2. If the game results in 4 tails followed by a head, player A pays player B $2^5 = \$32$. Therefore, the payoff to player B is a random variable that takes on the values 2^n for $n = 1, 2, \ldots$ and whose probability distribution is given by $(1/2)^n$ for $n = 1, 2, \ldots$, that is, if X denotes the payoff to player B,

$$P(X = 2^n) = \left(\frac{1}{2}\right)^n \text{ for } n = 1, 2, \ldots$$

The usual definition of a fair game between two players is for each player to have equal expectation for the amount to be won.
(a) How much should player B pay to player A so that this game will be fair?
(b) What is the variance of X?
(c) What is the probability of player B winning no more than \$8 in one play of the game?

24-12. The demand D for a product in a week is a random variable taking on the values of $-1, 0, 1$ with probabilities $1/8$, $5/8$, and $C/8$, respectively. A demand of -1 implies that an item is returned.
(a) Find C, $E(D)$, and variance D.
(b) Find $E(e^{D^2})$.
(c) Sketch the CDF of the random variable D, labeling all the necessary values.

24-13. In a certain chemical process three bottles of a standard fluid are emptied into a larger container. A study of the individual bottles shows that the mean value of the contents is 15 ounces and the standard deviation is 0.08 ounces. If three bottles form a random sample,
(a) Find the expected value and the standard deviation of the volume of liquid emptied into the larger container.
(b) If the content of the individual bottles is normally distributed, what is the probability that the volume of liquid emptied into the larger container will be in excess of 45.2 ounces?

24-14. Consider the density function of a random variable X defined by

$$f_X(y) = \begin{cases} 0, & \text{for } y < 0 \\ 6y(1 - y), & \text{for } 0 \leq y \leq 1 \\ 0, & \text{for } 1 < y. \end{cases}$$

(a) Find the CDF corresponding to this density function. (Be sure you describe it completely.)
(b) Calculate the mean and variance.
(c) What is the probability that a random variable having this density will exceed 0.5?
(d) Consider the experiment where six independent random variables are observed, each random variable having the density function given above. What is the expected value of the sample mean of these observations?
(e) What is the variance of the sample mean described in part (d)?

24-15. A transistor radio operates on two $1\frac{1}{2}$ volt batteries, so that nominally it operates on 3 volts. Suppose the actual voltage of a single new battery is normally distributed with mean $1\frac{1}{2}$volts and variance 0.0625. The radio will not operate "properly" at the outset if the voltage falls outside the range $2\frac{3}{4}$ to $3\frac{1}{4}$ volts.

(a) What is the probability that the radio will not operate "properly"?
(b) Suppose that the assumption of normality is not valid. Give a bound on the probability that the radio will not operate "properly."

24-16. The life of electric lightbulbs is known to be a normally distributed random variable with unknown mean μ and standard deviation 200 hours. The value of a lot of 1,000 bulbs is (1,000)(1/5,000) μ dollars. A random sample of n bulbs is to be drawn by a prospective buyer, and 1,000(1/5,000) $\overline{X}$ dollars paid to the manufacturer. How large should n be so that the probability is 0.90 that the buyer does not overpay or underpay the manufacturer by more than \$15?

24-17. A joint random variable (X_1, X_2) is said to have a bivariate normal distribution if its joint density is given by

$$f_{X_1, X_2}(s, t) = \frac{1}{2\pi\sigma_{X_1}\sigma_{X_2}\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{s-\mu_{X_1}}{\sigma_{X_1}}\right)^2 - 2\rho\frac{(s-\mu_{X_1})(t-\mu_{X_2})}{\sigma_{X_1}\sigma_{X_2}} + \left(\frac{t-\mu_{X_2}}{\sigma_{X_2}}\right)^2\right]\right\}$$

for $-\infty < s < \infty$ and $-\infty < t < \infty$.
(a) Show that $E(X_1) = \mu_{X_1}$ and $E(X_2) = \mu_{X_2}$.
(b) Show that variance $(X_1) = \sigma^2_{X_1}$, variance $(X_2) = \sigma^2_{X_2}$, and the correlation coefficient is ρ.
(c) Show that marginal distributions of X_1 and X_2 are normal.
(d) Show that the conditional distribution of X_1, given $X_2 = x_2$, is normal with mean

$$\mu_{X_1} + \rho\frac{\sigma_{X_1}}{\sigma_{X_2}}(x_2 - \mu_{X_2})$$

and variance $\sigma^2_{X_1}(1-\rho^2)$.

24-18. The joint demand for a product over 2 months is a continuous random variable (X_1, X_2) having a joint density given by

$$f_{X_1, X_2}(s, t) = \begin{cases} c, & \text{if } 100 \le s \le 150, \text{ and } 50 \le t \le 100 \\ 0, & \text{otherwise.} \end{cases}$$

(a) Find c.
(b) Find $F_{X_1X_2}(b_1, b_2)$, $F_{X_1}(b_1)$, and $F_{X_2}(b_2)$.
(c) Find $f_{X_2|X_1=s}(t)$.

24-19. Two machines produce a certain item. The capacity per day of machine 1 is 1 unit and that of machine 2 is 2 units. Let (X_1, X_2) be the discrete random variable that measures the actual production on each machine per day. Each entry in the table below represents the joint probability, for example, $P_{X_1X_2}(0,0) = {}^1/_8$.

X_2 \ X_1	0	1
0	$\frac{1}{8}$	0
1	$\frac{1}{4}$	$\frac{1}{8}$
2	$\frac{1}{8}$	$\frac{3}{8}$

(a) Find the marginal distributions of X_1 and X_2.
(b) Find the conditional distribution of X_1, given $X_2 = 1$.
(c) Are X_1 and X_2 independent random variables?
(d) Find $E(X_1)$, $E(X_2)$, variance (X_1), and variance (X_2).
(e) Find the probability distribution of $(X_1 + X_2)$.

24-20. Suppose that $E_1, E_2, \ldots, E_m$ are mutually exclusive events such that $E_1 \cup E_2 \cup \cdots \cup E_m = \Omega$; that is, exactly one of the E events will occur. Denote by F any event in the sample space. Note that

$$F = FE_1 \cup FE_2 \cup \cdots \cup FE_m\dagger$$

and that FE_1, $i = 1, 2, \ldots, m$, are also mutually exclusive.
(a) Show that $P\{F\} = \sum_{i=1}^{m} P\{FE_i\} = \sum_{i=1}^{m} P\{F|E_i\}P\{E_i\}$.
(b) Show that $P\{E_i|F\} = P\{F|E_i\}P\{E_i\}\Big/\sum_{i=1}^{m} P\{F|E_i\}P\{E_i\}$.
(This result is called Bayes' formula and is useful when it is known that the event F has occurred and there is interest in determining which one of the E_t also occurred.)

†Recall that FE_1 is the same as $F \cap E_1$, that is, the intersection of the two events F and E_1.