

---

## RANDOM VARIABLES AND THEIR DISTRIBUTIONS

---

### 2.1

#### INTRODUCTION

Our purpose is to develop mathematical models for describing the probabilities of outcomes or events occurring in a sample space. Because mathematical equations are expressed in terms of numerical values rather than as heads, colors, or other properties, it is convenient to define a function, known as a random variable, that associates each outcome in the experiment with a real number. We then can express the probability model for the experiment in terms of this associated random variable. Of course, in many experiments the results of interest already are numerical quantities, and in that case the natural function to use as the random variable would be the identity function.

**Definition 2.1.1**

**Random Variable** A random variable, say  $X$ , is a function defined over a sample space,  $S$ , that associates a real number,  $X(e) = x$ , with each possible outcome  $e$  in  $S$ .

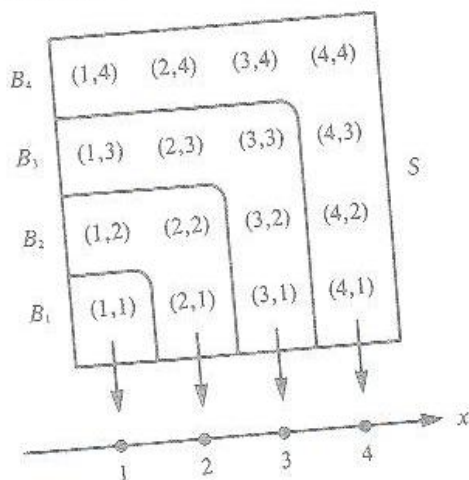
Capital letters, such as  $X$ ,  $Y$ , and  $Z$  will be used to denote random variables. The lower case letters  $x$ ,  $y$ ,  $z$ , ... will be used to denote possible values that the corresponding random variables can attain. For mathematical reasons, it will be necessary to restrict the types of functions that are considered to be random variables. We will discuss this point after the following example.

**Example 2.1.1**

A four-sided (tetrahedral) die has a different number—1, 2, 3, or 4—affixed to each side. On any given roll, each of the four numbers is equally likely to occur. A game consists of rolling the die twice, and the score is the *maximum* of the two numbers that occur. Although the score cannot be predicted, we can determine the set of possible values and define a random variable. In particular, if  $e = (i, j)$ , where  $i, j \in \{1, 2, 3, 4\}$ , then  $X(e) = \max(i, j)$ . The sample space,  $S$ , and  $X$  are illustrated in Figure 2.1.

**FIGURE 2.1**

Sample space for two rolls of a four-sided die



Each of the events  $B_1$ ,  $B_2$ ,  $B_3$ , and  $B_4$  of  $S$  contains the pairs  $(i, j)$  that have a common maximum. In other words,  $X$  has value  $x = 1$  over  $B_1$ ,  $x = 2$  over  $B_2$ ,  $x = 3$  over  $B_3$ , and  $x = 4$  over  $B_4$ .

Other random variables also could be considered. For example, the random variable  $Y(e) = i + j$  represents the total on the two rolls.

The concept of a random variable permits us to associate with any sample space,  $S$ , a sample space that is a set of real numbers, and in which the events of interest are subsets of real numbers. If such a real-valued event is denoted by  $A$ ,

## 2.1 INTRODUCTION

then we would want the associated set

$$B = \{e | e \in S \text{ and } X(e) \in A\} \quad (2.1.1)$$

to be an event in the underlying sample space  $S$ . Even though  $A$  and  $B$  are subsets of different spaces, they usually are referred to as **equivalent events**, and we write

$$P[X \in A] = P(B) \quad (2.1.2)$$

The notation  $P_X(A)$  sometimes is used instead of  $P[X \in A]$  in equation (2.1.2). This defines a set function on the collection of real-valued events, and it can be shown to satisfy the three basic conditions of a probability set function, as given by Definition 1.3.1.

Although the random variable  $X$  is defined as a function of  $e$ , it usually is possible to express the events of interest only in terms of the real values that  $X$  assumes. Thus, our notation usually will suppress the dependence on the outcomes in  $S$ , such as we have done in equation (2.1.2).

For instance, in Example 2.1.1, if we were interested in the event of obtaining a score of "at most 3," this would correspond to  $X = 1, 2, \text{ or } 3$ , or  $X \in \{1, 2, 3\}$ . Another possibility would be to represent the event in terms of some interval that contains the values 1, 2, and 3 but not 4, such as  $A = (-\infty, 3]$ . The associated equivalent event in  $S$  is  $B = B_1 \cup B_2 \cup B_3$ , and the probability is  $P[X \in A] = P(B) = 1/16 + 3/16 + 5/16 = 9/16$ . A convenient notation for  $P[X \in A]$ , in this example, is  $P[X \leq 3]$ . Actually, any other real event containing 1, 2, and 3 but not 4 could be used in this way, but intervals, and especially those of the form  $(-\infty, x]$ , will be of special importance in developing the properties of random variables.

As mentioned in Section 1.3, if the probabilities can be determined for each elementary event in a discrete sample space, then the probability of any event can be calculated from these by expressing the event as a union of mutually exclusive elementary events, and summing over their probabilities.

A more general approach for assigning probabilities to events in a real sample space can be based on assigning probabilities to intervals of the form  $(-\infty, x]$  for all real numbers  $x$ . Thus, we will consider as random variables only functions  $X$  that satisfy the requirements that, for all real  $x$ , sets of the form

$$B = [X \leq x] = \{e | e \in S \text{ and } X(e) \in (-\infty, x]\} \quad (2.1.3)$$

are events in the sample space  $S$ . The probabilities of other real events can be evaluated in terms of the probabilities assigned to such intervals. For example, for the game of Example 2.1.1, we have determined that  $P[X \leq 3] = 9/16$ , and it also follows, by a similar argument, that  $P[X \leq 2] = 1/4$ . Because  $(-\infty, 2]$  contains 1 and 2 but not 3, and  $(-\infty, 3] = (-\infty, 2] \cup (2, 3]$ , it follows that  $P[X = 3] = P[X \leq 3] - P[X \leq 2] = 9/16 - 1/4 = 5/16$ .

Other examples of random variables can be based on the sampling problems of Section 1.6.

**Example 2.1.2**

In Example 1.6.15, we discussed several alternative approaches for computing the probability of obtaining "exactly two black" marbles, when selecting five (without replacement) from a collection of 10 black and 20 white marbles. Suppose we are concerned with the general problem of obtaining  $x$  black marbles, for arbitrary  $x$ . Our approach will be to define a random variable  $X$  as the number of black marbles in the sample, and to determine the probability  $P[X = x]$  for every possible value  $x$ . This is easily accomplished with the approach given by equation (1.6.8), and the result is

$$P[X = x] = \frac{\binom{10}{x} \binom{20}{5-x}}{\binom{30}{5}} \quad x = 0, 1, 2, 3, 4, 5 \quad (2.1.4)$$

Random variables that arise from counting operations, such as the random variables in Examples 2.1.1. and 2.1.2, are integer-valued. Integer-valued random variables are examples of an important special type known as discrete random variables.

## 2.2

## DISCRETE RANDOM VARIABLES

**Definition 2.2.1**

If the set of all possible values of a random variable,  $X$ , is a countable set,  $x_1, x_2, \dots, x_n$ , or  $x_1, x_2, \dots$ , then  $X$  is called a **discrete random variable**. The function

$$f(x) = P[X = x] \quad x = x_1, x_2, \dots \quad (2.2.1)$$

that assigns the probability to each possible value  $x$  will be called the **discrete probability density function** (discrete pdf).

If it is clear from the context that  $X$  is discrete, then we simply will say pdf. Another common terminology is **probability mass function** (pmf), and the possible values,  $x_i$ , are called **mass points** of  $X$ . Sometimes a subscripted notation,  $f_X(x)$ , is used.

The following theorem gives general properties that any discrete pdf must satisfy.

**Theorem 2.2.1** A function  $f(x)$  is a discrete pdf if and only if it satisfies both of the following properties for at most a countably infinite set of reals  $x_1, x_2, \dots$ :

$$f(x_i) \geq 0 \quad (2.2.2)$$

for all  $x_i$ , and

$$\sum_{\text{all } x_i} f(x_i) = 1 \quad (2.2.3)$$

**Proof**

Property (2.2.2) follows from the fact that the value of a discrete pdf is a probability and must be nonnegative. Because  $x_1, x_2, \dots$  represent all possible values of  $X$ , the events  $[X = x_1], [X = x_2], \dots$  constitute an exhaustive partition of the sample space. Thus,

$$\sum_{\text{all } x_i} f(x_i) = \sum_{\text{all } x_i} P[X = x_i] = 1$$

Consequently, any pdf must satisfy properties (2.2.2) and (2.2.3) and any function that satisfies properties (2.2.2) and (2.2.3) will assign probabilities consistent with Definition 1.3.1. ■

In some problems, it is possible to express the pdf by means of an equation, such as equation (2.1.4). However, it is sometimes more convenient to express it in tabular form. For example, one way to specify the pdf of  $X$  for the random variable  $X$  in Example 2.1.1 is given in Table 2.1.

**TABLE 2.1**

Values of the discrete pdf  
of the maximum of two rolls  
of a four-sided die

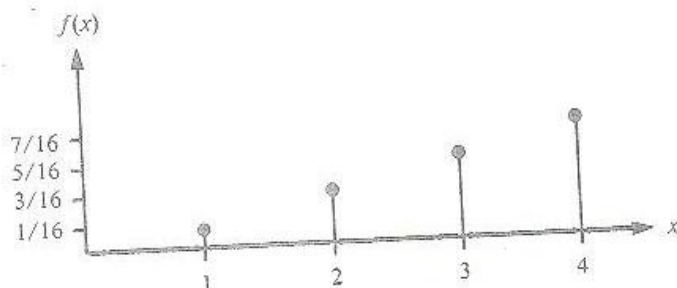
$x$	1	2	3	4
$f(x)$	1/16	3/16	5/16	7/16

Of course, these are the probabilities, respectively, of the events  $B_1, B_2, B_3$ , and  $B_4$  in  $S$ .

A graphic representation of  $f(x)$  is also of some interest. It would be possible to leave  $f(x)$  undefined at points that are not possible values of  $X$ , but it is convenient to define  $f(x)$  as zero at such points. The graph of the pdf in Table 2.1 is shown in Figure 2.2.

FIGURE 2.2

Discrete pdf of the maximum of two rolls of a four-sided die

**Example 2.2.1**

Example 2.1.1 involves two rolls of a four-sided die. Now we will roll a 12-sided (dodecahedral) die twice. If each face is marked with an integer, 1 through 12, then each value is equally likely to occur on a single roll of the die. As before, we define a random variable  $X$  to be the maximum obtained on the two rolls. It is not hard to see that for each value  $x$  there are an odd number,  $2x - 1$ , of ways for that value to occur. Thus, the pdf of  $X$  must have the form

$$f(x) = c(2x - 1) \quad \text{for } x = 1, 2, \dots, 12 \quad (2.2.4)$$

One way to determine  $c$  would be to do a more complete analysis of the counting problem, but another way would be to use equation (2.2.3). In particular,

$$\begin{aligned} 1 &= \sum_{x=1}^{12} f(x) = c \sum_{x=1}^{12} (2x - 1) = c \left[ 2 \sum_{x=1}^{12} x - 12 \right] \\ &= c \left[ \frac{2(12)(13)}{2} - 12 \right] = c(12)^2 \end{aligned}$$

So  $c = 1/(12)^2 = 1/144$ .

As mentioned in the last section, another way to specify the distribution of probability is to assign probabilities to intervals of the form  $(-\infty, x]$ , for all real  $x$ . The probability assigned to such an event is given by a function called the cumulative distribution function.

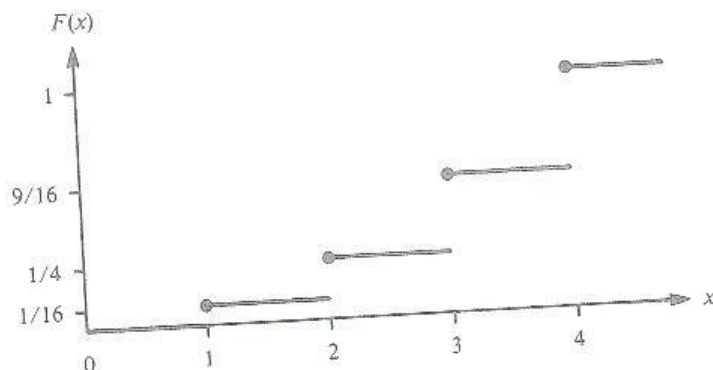
**Definition 2.2.2**

The cumulative distribution function (CDF) of a random variable  $X$  is defined for any real  $x$  by

$$F(x) = P[X \leq x] \quad (2.2.5)$$

FIGURE 2.3

The CDF of the maximum of two rolls of a four-sided die



The function  $F(x)$  often is referred to simply as the **distribution function** of  $X$ , and the subscripted notation,  $F_X(x)$ , sometimes is used.

For brevity, we often will use a short notation to indicate that a distribution of a particular form is appropriate. If we write  $X \sim f(x)$  or  $X \sim F(x)$ , this will mean that the random variable  $X$  has pdf  $f(x)$  and CDF  $F(x)$ .

As seen in Figure 2.3, the CDF of the distribution given in Table 2.1 is a nondecreasing step function. The step-function form of  $F(x)$  is common to all discrete distributions, and the sizes of the steps or jumps in the graph of  $F(x)$  correspond to the values of  $f(x)$  at those points. This is easily seen by comparing Figures 2.2 and 2.3.

The general relationship between  $F(x)$  and  $f(x)$  for a discrete distribution is given by the following theorem.

**Theorem 2.2.2** Let  $X$  be a discrete random variable with pdf  $f(x)$  and CDF  $F(x)$ . If the possible values of  $X$  are indexed in increasing order,  $x_1 < x_2 < x_3 < \dots$ , then  $f(x_1) = F(x_1)$ , and for any  $i > 1$ ,

$$f(x_i) = F(x_i) - F(x_{i-1}) \quad (2.2.6)$$

Furthermore, if  $x < x_1$  then  $F(x) = 0$ , and for any other real  $x$

$$F(x) = \sum_{x_i \leq x} f(x_i) \quad (2.2.7)$$

where the summation is taken over all indices  $i$  such that  $x_i \leq x$ . ■

The CDF of any random variable must satisfy the properties of the following theorem.

**Theorem 2.2.3** A function  $F(x)$  is a CDF for some random variable  $X$  if and only if it satisfies the following properties:

$$\lim_{x \rightarrow -\infty} F(x) = 0 \quad (2.2.8)$$

$$\lim_{x \rightarrow \infty} F(x) = 1 \quad (2.2.9)$$

$$\lim_{h \rightarrow 0^+} F(x+h) = F(x) \quad (2.2.10)$$

$$a < b \text{ implies } F(a) \leq F(b) \quad (2.2.11)$$

The first two properties say that  $F(x)$  can be made arbitrarily close to 0 or 1 by taking  $x$  arbitrarily large, and negative or positive, respectively. In the examples considered so far, it turns out that  $F(x)$  actually assumes these limiting values. Property (2.2.10) says that  $F(x)$  is *continuous from the right*. Notice that in Figure 2.3 the only discontinuities are at the values 1, 2, 3, and 4, and the limit as  $x$  approaches these values from the right is the value of  $F(x)$  at these values. On the other hand, as  $x$  approaches these values from the left, the limit of  $F(x)$  is the value of  $F(x)$  on the lower step, so  $F(x)$  is not (in general) continuous from the left. Property (2.2.11) says that  $F(x)$  is *nondecreasing*, which is easily seen to be the case in Figure 2.3. In general, this property follows from the fact that an interval of the form  $(-\infty, b]$  can be represented as the union of two disjoint intervals

$$(-\infty, b] = (-\infty, a] \cup (a, b] \quad (2.2.12)$$

for any  $a < b$ . It follows that  $F(b) = F(a) + P[a < x \leq b] \geq F(a)$ , because  $P[a < x \leq b] \geq 0$ , and thus equation (2.2.11) is obtained.

Actually, by this argument we have obtained another very useful result, namely,

$$P[a < X \leq b] = F(b) - F(a) \quad (2.2.13)$$

This reduces the problem of computing probabilities for events defined in terms of intervals of the form  $(a, b]$  to taking differences with  $F(x)$ .

Generally, it is somewhat easier to understand the nature of a random variable and its probability distribution by considering the pdf directly, rather than the CDF, although the CDF will provide a good basis for defining continuous probability distributions. This will be considered in the next section.

Some important properties of probability distributions involve numerical quantities called expected values.

**Definition 2.2.3**

If  $X$  is a discrete random variable with pdf  $f(x)$ , then the expected value of  $X$  is defined by

$$E(X) = \sum_x x f(x) \quad (2.2.14)$$

The sum (2.2.14) is understood to be over all possible values of  $X$ . Furthermore, it is an ordinary sum if the range of  $X$  is finite, and an infinite series if the range of  $X$  is infinite. In the latter case, if the infinite series is not absolutely convergent, then we will say that  $E(X)$  does not exist. Other common notations for  $E(X)$  include  $\mu$ , possibly with a subscript,  $\mu_X$ . The terms **mean** and **expectation** also are often used.

The mean or expected value of a random variable is a "weighted average," and it can be considered as a measure of the "center" of the associated probability distribution.

**Example 2.2.2**

A box contains four chips. Two are labeled with the number 2, one is labeled with a 4, and the other with an 8. The average of the numbers on the four chips is  $(2 + 2 + 4 + 8)/4 = 4$ . The experiment of choosing a chip at random and recording its number can be associated with a discrete random variable  $X$  having distinct values  $x = 2, 4$ , or  $8$ , with  $f(2) = 1/2$  and  $f(4) = f(8) = 1/4$ . The corresponding expected value or mean is

$$\mu = E(X) = 2\left(\frac{1}{2}\right) + 4\left(\frac{1}{4}\right) + 8\left(\frac{1}{4}\right) = 4$$

as before. Notice that this also could model selection from a larger collection, as long as the possible observed values of  $X$  and the respective proportions in the collection,  $f(x)$ , remain the same as in the present example.

There is an analogy between the distribution of probability to values,  $x$ , and the distribution of mass to points in a physical system. For example, if masses of 0.5, 0.25, and 0.25 grams are placed at the respective points  $x = 2, 4$ , and  $8$  cm on the horizontal axis, then the value  $2(0.5) + 4(0.25) + 8(0.25) = 4$  is the "center of mass" or balance point of the corresponding system. This is illustrated in Figure 2.4.

**FIGURE 2.4**

The center-of-mass interpretation of the mean



In the previous example  $E(X)$  coincides with one of the possible values of  $X$ , but this is not always the case, as illustrated by the following example.

**Example 2.2.3**

A game of chance is based on drawing two chips at random without replacement from the box considered in Example 2.2.2. If the numbers on the two chips match, then the player wins \$2; otherwise, she loses \$1. Let  $X$  be the amount won by the player on a single play of the game. There are only two possible values,  $X = 2$  if both chips bear the number 2, and  $X = -1$  otherwise. Furthermore, there are  $\binom{4}{2} = 6$  ways to draw two chips, and only one of these outcomes correspond to a match. The distribution of  $X$  is  $f(2) = 1/6$  and  $f(-1) = 5/6$ , and consequently the expected amount won is  $E(X) = (-1)(5/6) + (2)(1/6) = -1/2$ . Thus, the expected amount "won" by the player is actually an expected loss of one-half dollar.

The connection with long-term relative frequency also is well illustrated by this example. Suppose the game is played  $M$  times in succession, and denote the relative frequencies of winning and losing by  $f_w$  and  $f_L$ , respectively. The average amount the player wins is  $(-1)f_L + (2)f_w$ . Because of statistical regularity, we have that  $f_L$  and  $f_w$  approach  $f(-1)$  and  $f(2)$ , respectively, and thus the player's average winnings approach  $E(X)$  as  $M$  approaches infinity.

Notice also that the game will be more equitable if the payoff to the player is changed to \$5 rather than \$2, because the resulting expected amount won then will be  $(-1)(5/6) + (5)(1/6) = 0$ . In general, for a game of chance, if the net amount won by a player is  $X$ , then the game is said to be a *fair game* if  $E(X) = 0$ .

## 2.3

### CONTINUOUS RANDOM VARIABLES

The notion of a discrete random variable provides an adequate means of probability modeling for a large class of problems, including those that arise from the operation of counting. However, a discrete random variable is not an adequate model in many situations, and we must consider the notion of a continuous random variable. The CDF defined earlier remains meaningful for continuous random variables, but it also is useful to extend the concept of a pdf to continuous random variables.

**Example 2.3.1**

Each work day a man rides a bus to his place of business. Although a new bus arrives promptly every five minutes, the man generally arrives at the bus stop at a random time between bus arrivals. Thus, we might take his waiting time on any given morning to be a random variable  $X$ .

Although in practice we usually measure time only to the nearest unit (seconds, minutes, etc.), in theory we could measure time to within some arbitrarily small unit. Thus, even though in practice it might be possible to regard  $X$  as a discrete

random variable with possible values determined by the smallest appropriate time unit, it usually is more convenient to consider the idealized situation in which  $X$  is assumed capable of attaining any value in some interval, and not just discrete points.

Returning to the man waiting for his bus, suppose that he is very observant and noticed over the years that the frequency of days when he waits no more than  $x$  minutes for the bus is proportional to  $x$  for all  $x$ . This suggests a CDF of the form  $F(x) = P[X \leq x] = cx$ , for some constant  $c > 0$ . Because the buses arrive at regular five-minute intervals, the range of possible values of  $X$  is the time interval  $[0, 5]$ . In other words,  $P[0 \leq X \leq 5] = 1$ , and it follows that  $1 = F(5) = c \cdot 5$ , and thus  $c = 1/5$ , and  $F(x) = x/5$  if  $0 \leq x \leq 5$ . It also follows that  $F(x) = 0$  if  $x < 0$  and  $F(x) = 1$  if  $x > 5$ .

Another way to study this distribution would be to observe the relative frequency of bus arrivals during short time intervals of the same length, but distributed throughout the waiting-time interval  $[0, 5]$ . It may be that the frequency of bus arrivals during intervals of the form  $(x, x + \Delta x]$  for small  $\Delta x$  was proportional to the length of the interval,  $\Delta x$ , regardless of the value of  $x$ . The corresponding condition this imposes on the distribution of  $X$  is

$$P[x < X \leq x + \Delta x] = F(x + \Delta x) - F(x) = c \Delta x$$

for all  $0 \leq x < x + \Delta x \leq 5$  and some  $c > 0$ . Of course, this implies that if  $F(x)$  is differentiable at  $x$ , its derivative is constant,  $F'(x) = c > 0$ . Note also that for  $x < 0$  or  $x > 5$ , the derivative also exists, but  $F'(x) = 0$  because  $P[x < X \leq x + \Delta x] = 0$  when  $x$  and  $x + \Delta x$  are not possible values of  $X$ , and the derivative does not exist at all at  $x = 0$  or  $5$ .

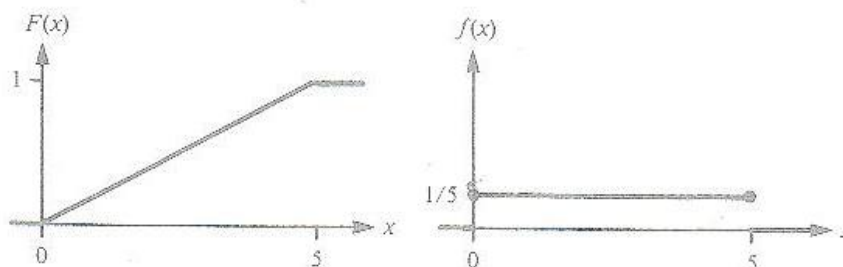
In general, if  $F(x)$  is the CDF of a continuous random variable  $X$ , then we will denote its derivative (where it exists) by  $f(x)$ , and under certain conditions, which will be specified shortly, we will call  $f(x)$  the probability density function of  $X$ . In our example,  $F(x)$  can be represented for values of  $x$  in the interval  $[0, 5]$  as the integral of its derivative:

$$F(x) = \int_{-\infty}^x f(t) dt = \int_0^x \frac{1}{5} dt = \frac{x}{5}$$

The graphs of  $F(x)$  and  $f(x)$  are shown in Figure 2.5.

FIGURE 2.5

CDF and pdf of waiting time for a bus



This provides a general approach to defining the distribution of a continuous random variable  $X$ .

**Definition 2.3.1**

A random variable  $X$  is called a continuous random variable if there is a function  $f(x)$ , called the probability density function (pdf) of  $X$ , such that the CDF can be represented as

$$F(x) = \int_{-\infty}^x f(t) dt \quad (2.3.1)$$

In more advanced treatments of probability, such distributions sometimes are called "absolutely continuous" distributions. The reason for such a distinction is that CDFs exist that are continuous (in the usual sense), but which cannot be represented as the integral of the derivative. We will apply the terminology continuous distribution only to probability distributions that satisfy property (2.3.1).

Sometimes it is convenient to use a subscripted notation,  $F_X(x)$  and  $f_X(x)$ , for the CDF and pdf, respectively.

The defining property (2.3.1) provides a way to derive the CDF when the pdf is given, and it follows by the Fundamental Theorem of Calculus that the pdf can be obtained from the CDF by differentiation. Specifically,

$$f(x) = \frac{d}{dx} F(x) = F'(x) \quad (2.3.2)$$

wherever the derivative exists. Recall from Example 2.3.1 that there were two values of  $x$  where the derivative of  $F(x)$  did not exist. In general, there may be many values of  $x$  where  $F(x)$  is not differentiable, and these will occur at discontinuity points of the pdf,  $f(x)$ . Inspection of the graphs of  $f(x)$  and  $F(x)$  in Figure 2.5 shows that this situation occurs in the example at  $x = 0$  and  $x = 5$ . However, this will not usually create a problem if the set of such values is finite, because an integrand can be redefined arbitrarily at a finite number of values  $x$  without affecting the value of the integral. Thus, the function  $F(x)$ , as represented in property (2.3.1), is unaffected regardless of how we treat such values. It also follows by similar considerations that events such as  $[X = c]$ , where  $c$  is a constant, will have probability zero when  $X$  is a continuous random variable. Consequently, events of the form  $[X \in I]$ , where  $I$  is an interval, are assigned the same probability whether  $I$  includes the endpoints or not. In other words, for a continuous random variable  $X$ , if  $a < b$ ,

$$\begin{aligned} P[a < X \leq b] &= P[a \leq X < b] = P[a < X < b] \\ &= P[a \leq X \leq b] \end{aligned} \quad (2.3.3)$$

and each of these has the value  $F(b) - F(a)$ .

Thus, the CDF,  $F(x)$ , assigns probabilities to events of the form  $(-\infty, x]$ , and equation (2.3.3) shows how the probability assignment can be extended to any interval.

Any function  $f(x)$  may be considered as a possible candidate for a pdf if it produces a legitimate CDF when integrated as in property (2.3.1). The following theorem provides conditions that will guarantee this.

**Theorem 2.3.1** A function  $f(x)$  is a pdf for some continuous random variable  $X$  if and only if it satisfies the properties

$$f(x) \geq 0 \quad (2.3.4)$$

for all real  $x$ , and

$$\int_{-\infty}^{\infty} f(x) dx = 1 \quad (2.3.5)$$

**Proof**

Properties (2.2.9) and (2.2.11) of a CDF follow from properties (2.3.5) and (2.3.4), respectively. The other properties follow from general results about integrals. ■

**Example 2.3.2** A machine produces copper wire, and occasionally there is a flaw at some point along the wire. The length of wire (in meters) produced between successive flaws is a continuous random variable  $X$  with pdf of the form

$$f(x) = \begin{cases} c(1+x)^{-3} & x > 0 \\ 0 & x \leq 0 \end{cases} \quad (2.3.6)$$

where  $c$  is a constant. The value of  $c$  can be determined by means of property (2.3.5). Specifically, set

$$1 = \int_{-\infty}^{\infty} f(x) dx = \int_0^{\infty} c(1+x)^{-3} dx = c \left( \frac{1}{2} \right)$$

which is obtained following the substitution  $u = 1 + x$  and an application of the power rule for integrals. This implies that the constant is  $c = 2$ .

Clearly property (2.3.4) also is satisfied in this case.

The CDF for this random variable is given by

$$\begin{aligned}
 F(x) &= P[X \leq x] = \int_{-\infty}^x f(t) dt \\
 &= \begin{cases} \int_{-\infty}^0 0 dt + \int_0^x 2(1+t)^{-3} dt & x > 0 \\ \int_{-\infty}^x 0 dt & x \leq 0 \end{cases} \\
 &= \begin{cases} 1 - (1+x)^{-2} & x > 0 \\ 0 & x \leq 0 \end{cases}
 \end{aligned}$$

Probabilities of intervals, such as  $P[a \leq X \leq b]$ , can be expressed directly in terms of the CDF or as integrals of the pdf. For example, the probability that a flaw occurs between 0.40 and 0.45 meters is given by

$$P[0.40 \leq X \leq 0.45] = \int_{0.40}^{0.45} f(x) dx = F(0.45) - F(0.40) = 0.035$$

Consideration of the frequency of occurrences over short intervals was suggested as a possible way to study a continuous distribution in Example 2.3.1. This approach provides some insight into the general nature of continuous distributions. For example, it may be observed that the frequency of occurrences over short intervals of length  $\Delta x$ , say  $[x, x + \Delta x]$ , is at least approximately proportional to the length of the interval,  $\Delta x$ , where the proportionality factor depends on  $x$ , say  $f(x)$ . The condition this imposes on the distribution of  $X$  is

$$\begin{aligned}
 P[x \leq X \leq x + \Delta x] &= F(x + \Delta x) - F(x) \\
 &\approx f(x) \Delta x
 \end{aligned} \tag{2.3.7}$$

where the error in the approximation is negligible relative to the length of the interval,  $\Delta x$ . This is illustrated in Figure 2.6. for the copper wire example.

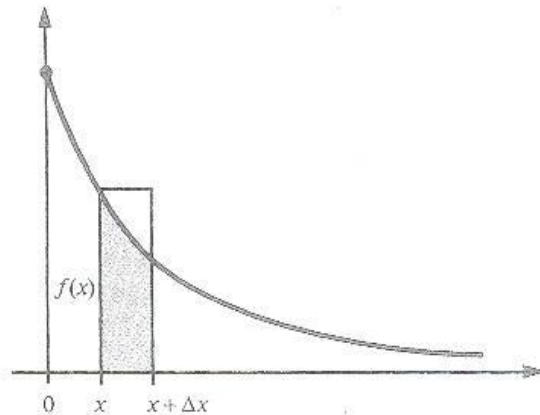
The exact probability in equation (2.3.7) is represented by the area of the shaded region under the graph of  $f(x)$ , while the approximation is the area of the corresponding rectangle with height  $f(x)$  and width  $\Delta x$ .

The smaller the value of  $\Delta x$ , the closer this approximation becomes. In this sense, it might be reasonable to think of  $f(x)$  as assigning "probability density" for the distribution of  $X$ , and the term probability density function seems appropriate for  $f(x)$ . In other words, for a continuous random variable  $X$ ,  $f(x)$  is not a probability, although it does determine the probability assigned to arbitrarily small intervals. The area between the  $x$ -axis and the graph of  $f(x)$  assigns probability to intervals, so that for  $a < b$ ,

$$P[a \leq X \leq b] = \int_a^b f(x) dx \tag{2.3.8}$$

FIGURE 2.6

Continuous assignment of probability by pdf



In Example 2.3.2, we could take the probability that the length between successive flaws between 0.40 and 0.45 meters to be approximately  $f(0.40)(0.05) = 2(1.4)^{-3}(0.05) = 0.036$ , or we could integrate the pdf between the limits 0.40 and 0.45 to obtain the exact answer, 0.035. For longer intervals, integrating  $f(x)$  as in equation (2.3.8) would be more reasonable.

Note that in Section 2.2 we referred to a probability density function or density function for a discrete random variable, but the interpretation there is different, because probability is assigned at discrete points in that case rather than in a continuous manner. However, it will be convenient to refer to the “density function” or pdf in both continuous and discrete cases, and to use the same notation,  $f(x)$  or  $f_X(x)$ , in the later chapters of the book. This will avoid the necessity of separate statements of general results that apply to both cases.

The notion of expected value can be extended to continuous random variables.

#### Definition 2.3.2

If  $X$  is a continuous random variable with pdf  $f(x)$ , then the expected value of  $X$  is defined by

$$E(X) = \int_{-\infty}^{\infty} xf(x) dx \quad (2.3.9)$$

if the integral in equation (2.3.9) is absolutely convergent. Otherwise we say that  $E(X)$  does not exist.

As in the discrete case, other notations for  $E(X)$  are  $\mu$  or  $\mu_X$ , and the terms mean or expectation of  $X$  also are commonly used. The center-of-mass analogy is

still valid in this case, where mass is assigned to the  $x$ -axis in a continuous manner and in accordance with  $f(x)$ . Thus,  $\mu$  can also be regarded as a central measure for a continuous distribution.

In Example 2.3.2, the mean length between flaws in a piece of wire is

$$\mu = \int_{-\infty}^0 x \cdot 0 \, dx + \int_0^{\infty} x \cdot 2(1+x)^{-3} \, dx$$

If we make the substitution  $t = 1 + x$ , then

$$\mu = 2 \int_1^{\infty} (t-1)t^{-3} \, dt = 2 \left( 1 - \frac{1}{2} \right) = 1$$

Other properties of probability distributions can be described in terms of quantities called percentiles.

**Definition 2.3.3**

If  $0 < p < 1$ , then a  $100 \times p$ th percentile of the distribution of a continuous random variable  $X$  is a solution  $x_p$  to the equation

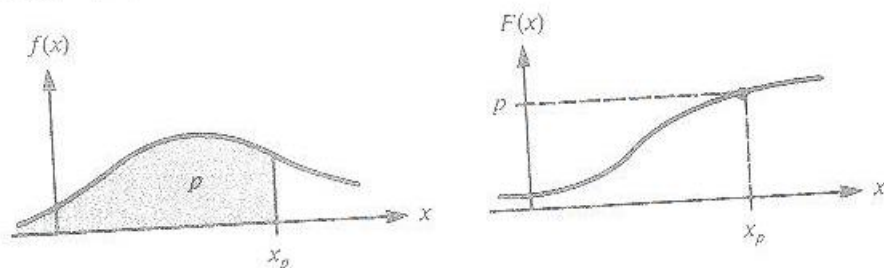
$$F(x_p) = p \quad (2.3.10)$$

In general, a distribution may not be continuous, and if it has a discontinuity, then there will be some values of  $p$  for which equation (2.3.10) has no solution. Although we emphasize the continuous case in this book, it is possible to state a general definition of percentile by defining a  $p$ th percentile of the distribution of  $X$  to be a value  $x_p$  such that  $P[X \leq x_p] \geq p$  and  $P[X \geq x_p] \geq 1 - p$ .

In essence,  $x_p$  is a value such that  $100 \times p$  percent of the population values are at most  $x_p$  and  $100 \times (1 - p)$  percent of the population values are at least  $x_p$ . This is illustrated for a continuous distribution in Figure 2.7. We also can think in terms of a proportion  $p$  rather than a percentage  $100 \times p$  of the population, and in this context  $x_p$  is called a  $p$ th quantile of the distribution.

FIGURE 2.7

A  $100 \times p$ th percentile



A **median** of the distribution of  $X$  is a 50th percentile, denoted by  $x_{0.5}$  or  $m$ . This is an important special case of the percentile such that half of the population values are above it and half are below it. The median is used in some applications instead of the mean as a central measure.

**Example 2.3.3**

Consider the distribution of lifetimes,  $X$  (in months), of a particular type of component. We will assume that the CDF has the form

$$F(x) = 1 - e^{-(x/3)^2} \quad x > 0$$

and zero otherwise. The median lifetime is

$$m = 3[-\ln(1 - 0.5)]^{1/2} = 3\sqrt{\ln 2} = 2.498 \text{ months}$$

It is desired to find the time  $t$  such that 10% of the components fail before  $t$ . This is the 10th percentile:

$$x_{0.10} = 3[-\ln(1 - 0.1)]^{1/2} = 3\sqrt{-\ln(0.9)} = 0.974 \text{ months}$$

Thus, if the components are guaranteed for one month, slightly more than 10% will need to be replaced.

Another measure of central tendency, which is sometimes considered, is the **mode**.

**Definition 2.3.4**

If the pdf has a unique maximum at  $x = m_0$ , say  $\max f(x) = f(m_0)$ , then  $m_0$  is called the **mode** of  $X$ .

In the previous example, the pdf of the distribution of lifetimes is

$$f(x) = \left(\frac{2}{9}\right)xe^{-(x/3)^2} \quad x > 0$$

The solution to  $f'(x) = 0$  is the unique maximum of  $f(x)$ ,  $x = m_0 = 3\sqrt{2}/2 = 2.121$  months.

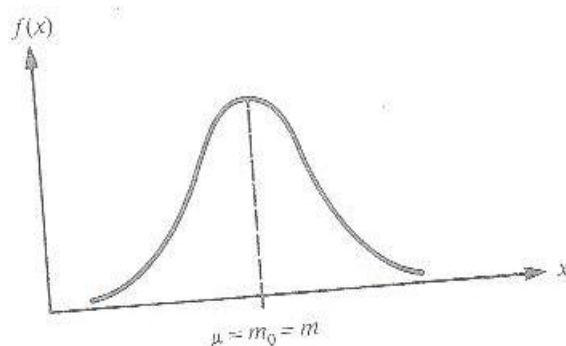
In general, the mean, median, and mode may be all different, but there are cases in which they all agree.

**Definition 2.3.5**

A distribution with pdf  $f(x)$  is said to be **symmetric** about  $c$  if  $f(c - x) = f(c + x)$  for all  $x$ .

FIGURE 2.8

The pdf of a symmetric distribution



In other words, the “centered” pdf  $g(x) = f(c - x)$  is an even function, in the usual sense that  $g(x) = g(-x)$ . The graph of  $y = f(x)$  is a “mirror image” about the vertical line  $x = c$ . Asymmetric distributions, such as the one in Example 2.3.2, are called **skewed** distributions.

If  $f(x)$  is symmetric about  $c$  and the mean  $\mu$  exists, then  $c = \mu$ . If additionally,  $f(x)$  has a unique maximum at  $m_0$  and a unique median  $m$ , then  $\mu = m_0 = m$ . This is illustrated in Figure 2.8.

### MIXED DISTRIBUTIONS

It is possible to have a random variable whose distribution is neither purely discrete nor continuous. A probability distribution for a random variable  $X$  is of **mixed type** if the CDF has the form

$$F(x) = aF_d(x) + (1 - a)F_c(x)$$

where  $F_d(x)$  and  $F_c(x)$  are CDFs of discrete and continuous type, respectively, and  $0 < a < 1$ .

#### Example 2.3.4

Suppose that a driver encounters a stop sign and either waits for a random period of time before proceeding or proceeds immediately. An appropriate model would allow the waiting time to be either zero or positive, both with nonzero probability. Let the CDF of the waiting time  $X$  be

$$\begin{aligned} F(x) &= 0.4F_d(x) + 0.6F_c(x) \\ &= 0.4 + 0.6(1 - e^{-x}) \end{aligned}$$

where  $F_d(x) = 1$  and  $F_c(x) = 1 - e^{-x}$  if  $x \geq 0$ , and both are zero if  $x < 0$ . The graph of  $F(x)$  is shown in Figure 2.9. Thus, the probability of proceeding immediately is  $P[X = 0] = 0.4$ . The probability that the waiting time is less than 0.5