4

One-Dimensional Random Variables

4.1 General Notion of a Random Variable

In describing the sample space of an experiment we did not specify that an individual outcome needs to be a number. In fact, we have cited a number of examples in which the result of the experiment was not a numerical quantity. For instance, in classifying a manufactured item we might simply use the categories "defective" and "nondefective." Again, in observing the temperature during a 24-hour period we might simply keep a record of the curve traced by the thermograph. However, in many experimental situations we are going to be concerned with measuring something and recording it as a *number*. Even in the above-cited cases we can assign a number to each (nonnumerical) outcome of the experiment. For example, we could assign the value one to nondefective items and the value zero to defective ones. We could record the maximum temperature of the day, or the minimum temperature, or the average of the maximum and minimum temperatures.

The above illustrations are quite typical of a very general class of problems: In many experimental situations we want to assign a real number x to every element s of the sample space S. That is, $x = X(s)$ is the value of a function X from the sample space to the real numbers. With this in mind, we make the following formal definition.

Definition. Let $\mathcal{E}$ be an experiment and S a sample space associated with the experiment. A *function* X assigning to every element $s \in S$, a real number, $X(s)$, is called a *random variable*.

Notes: (a) The above terminology is a somewhat unfortunate one, but it is so universally accepted that we shall not deviate from it. We have made it as clear as possible that X is a *function* and yet we call it a (random) variable!

(b) It turns out that *not every* conceivable function may be considered as a random variable. One requirement (although not the most general one) is that for every real number x the event $\{X(s) = x\}$ and for every interval I the event $\{X(s) \in I\}$ have well-defined probabilities consistent with the basic axioms. In most of the applications this difficulty does not arise and we shall make no further reference to it.

(c) In some situations the outcome s of the sample space is already the numerical characteristic which we want to record. We simply take $X(s) = s$, the identity function.

(d) In most of our subsequent discussions of random variables we need not indicate the functional nature of X. We are usually interested in the possible *values* of X, rather than where these values came from. For example, suppose that we toss two coins and consider the sample space associated with this experiment. That is,

$$S = \{HH, HT, TH, TT\}.$$

Define the random variable X as follows: X is the number of heads obtained in the two tosses. Hence $X(HH) = 2$, $X(HT) = X(TH) = 1$, and $X(TT) = 0$.

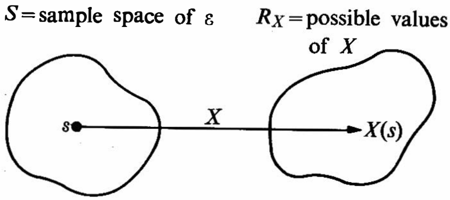

FIGURE 4.1

(e) It is very important to understand a basic requirement of a (single-valued) function: To *every* $s \in S$ there corresponds *exactly one* value $X(s)$. This is shown schematically in Fig. 4.1. Different values of s *may* lead to the same value of X. For example, in the above illustration we found that $X(HT) = X(TH) = 1$.

The space R_X, the set of all possible values of X, is sometimes called the *range space*. In a sense we may consider R_X as another sample space. The (original) sample space S corresponds to the (possibly) nonnumerical outcome of the experiment, while R_X is the sample space associated with the random variable X, representing the numerical characteristic which may be of interest. If $X(s) = s$, we have $S = R_X$.

Although we are aware of the pedagogical danger inherent in giving too many explanations for the same thing, let us nevertheless point out that we may think of a random variable X in two ways:

(a) We perform the experiment $\mathcal{E}$ which results in an outcome $s \in S$. We *then* evaluate the number $X(s)$.

(b) We perform $\mathcal{E}$, obtaining the outcome s, *and* (immediately) evaluate $X(s)$. The number $X(s)$ is then thought of as the actual outcome of the experiment and R_X becomes *the* sample space of the experiment.

The difference between interpretations (a) and (b) is hardly discernible. It is relatively minor but worthy of attention. In (a) the experiment essentially terminates with the observation of s. The evaluation of $X(s)$ is considered as something that is done subsequently and which is not affected by the randomness of $\mathcal{E}$. In (b) the experiment is not considered to be terminated until the number $X(s)$ has actually been evaluated, thus resulting in the sample space R_X. Although the first interpretation, (a), is the one usually intended, the second point of view, (b), can be very helpful, and the reader should keep it in mind. What we are saying, and this will become increasingly evident in later sections, is that in studying random variables we are more concerned about the values X assumes than about its functional form. Hence in many cases we shall completely ignore the underlying sample space on which X may be defined.

EXAMPLE 4.1. Suppose that a light bulb is inserted into a socket. The experiment is considered at an end when the bulb ceases to burn. What is a possible outcome, say s? One way of describing s would be by simply recording the date and time of day at which the bulb burns out, for instance May 19, 4:32 p.m. Hence the sample space may be represented as $S = \{(d, t) \mid d = \text{date}, t = \text{time of day}\}$. Presumably the random variable of interest is X, the length of burning time. Note that once $s = (d, t)$ is observed, the *evaluation* of $X(s)$ does not involve any randomness. When s is specified, $X(s)$ is completely determined.

The two points of view expressed above may be applied to this example as follows. In (a) we consider the experiment to be terminated with the observation $s = (d, t)$, the date and time of day. The computation of $X(s)$ is then performed, involving a simple arithmetic operation. In (b) we consider the experiment to be completed only *after* $X(s)$ is evaluated and the number $X(s) = 107$ hours, say, is then considered to be the outcome of the experiment.

It might be pointed out that a similar analysis could be applied to some other random variables of interest, for instance $Y(s)$ is the temperature in the room at the time the bulb burned out.

EXAMPLE 4.2. Three coins are tossed on a table. As soon as the coins land on the table, the "random" phase of the experiment is over. A single outcome s might consist of a detailed description of how and where the coins landed. Presumably we are only interested in certain numerical characteristics associated with this experiment. For instance, we might evaluate

$$X(s) = \text{number of heads showing,}$$
$$Y(s) = \text{maximum distance between any two coins,}$$
$$Z(s) = \text{minimum distance of coins from any edge of the table.}$$

If the random variable X is of interest, we could, as discussed in the previous example, incorporate the evaluation of $X(s)$ into the description of our experiment and hence simply state that the sample space associated with the experiment is $\{0, 1, 2, 3\}$, corresponding to the values of X. Although we shall very often adopt precisely this point of view, it is important to realize that the counting of the number of heads is done *after* the random aspects of the experiment have ended.

Note: In referring to random variables we shall, almost without exception, use capital letters such as X, Y, Z, etc. However, when speaking of the *value* these random variables assume we shall in general use lower case letters such as x, y, z, etc. This is a *very important distinction* to be made and the student might well pause to consider it. For example, when we speak of choosing a person at random from some designated population and measuring his height (in inches, say), we could refer to the *possible* outcomes as a random variable X. We might then ask various questions about X, such as $P(X \geq 60)$. However, once we actually choose a person and measure his height we obtain a specific value of X, say x. Thus it would be meaningless to ask for $P(x \geq 60)$ since x either is or is not ≥ 60. This distinction between a random variable and its value is important, and we shall make subsequent references to it.

As we were concerned about the events associated with the sample space S, so we shall find the need to discuss events with respect to the random variable X, that is, subsets of the range space R_X. Quite often certain events associated with S are "related" (in a sense to be described) to events associated with R_X in the following way.

Definition. Let ε be an experiment and S its sample space. Let X be a random variable defined on S and let R_X be its range space. Let B be an event with respect to R_X; that is, $B \subset R_X$. Suppose that A is defined as

$$A = \{s \in S \mid X(s) \in B\}. \tag{4.1}$$

In words: A consists of all outcomes in S for which $X(s) \in B$ (Fig. 4.2). In this case we say that A and B are *equivalent events.*

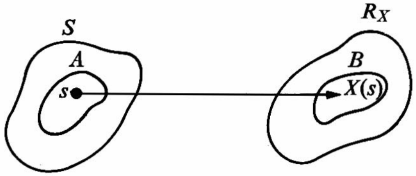

FIGURE 4.2

Notes: (a) Saying the above more informally, A and B are equivalent events whenever they occur together. That is, whenever A occurs, B occurs and conversely. For if A did occur, then an outcome s occurred for which $X(s) \in B$ and hence B occurred. Conversely, if B occurred, a value $X(s)$ was observed for which $s \in A$ and hence A occurred.

(b) It is important to realize that in our definition of equivalent events, A and B are associated with *different* sample spaces.

EXAMPLE 4.3. Consider the tossing of two coins. Hence $S = \{HH, HT, TH, TT\}$. Let X be the number of heads obtained. Hence $R_X = \{0, 1, 2\}$. Let $B = \{1\}$. Since $X(HT) = X(TH) = 1$ if and only if $X(s) = 1$, we have that $A = \{HT, TH\}$ is equivalent to B.

We now make the following important definition.

Definition. Let B be an event in the range space R_X. We then *define* $P(B)$ as follows:

$$P(B) = P(A), \quad \text{where} \quad A = \{s \in S \mid X(s) \in B\}. \tag{4.2}$$

In words: We define $P(B)$ equal to the probability of the event $A \subset S$, which is equivalent to B, in the sense of Eq. (4.1).

Notes: (a) We are assuming that probabilities may be associated with events in S. Hence the above definition makes it possible to assign probabilities to events associated with R_X in terms of probabilities defined over S.

(b) It is actually possible to *prove* that $P(B)$ must be as we defined it. However, this would involve some theoretical difficulties which we want to avoid, and hence we proceed as above.

(c) Since in the formulation of Eq. (4.2) the events A and B refer to different sample spaces, we should really use a different notation when referring to probabilities defined over S and for those defined over R_X, say something like $P(A)$ and $P_X(B)$. However, we shall not do this but continue to write simply $P(A)$ and $P(B)$. The context in which these expressions appear should make the interpretation clear.

(d) The probabilities associated with events in the (original) sample space S are, in a sense, determined by "forces beyond our control" or, as it is sometimes put, "by nature." The makeup of a radioactive source emitting particles, the disposition of a large number of persons who might place a telephone call during a certain hour, and the thermal agitation resulting in a current or the atmospheric conditions giving rise to a storm front illustrate this point. When we introduce a random variable X and its associated range space R_X we are *inducing* probabilities on the events associated with R_X, which are strictly determined if the probabilities associated with events in S are specified.

EXAMPLE 4.4. If the coins considered in Example 4.3 are "fair," we have $P(HT) = P(TH) = \frac{1}{4}$. Hence $P(HT, TH) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$. (The above calculations are a direct consequence of our basic assumption concerning the fairness of the coins.) Since the event $\{X = 1\}$ is equivalent to the event $\{HT, TH\}$, using Eq. (4.1), we have that $P(X = 1) = P(HT, TH) = \frac{1}{2}$. [There was really no choice about the value of $P(X = 1)$ consistent with Eq. (4.2), once $P(HT, TH)$ had been determined. It is in this sense that probabilities associated with events of R_X are *induced*.]

Note: Now that we have established the existence of an induced probability function over the range space of X (Eqs. 4.1 and 4.2) we shall find it convenient to *suppress* the functional nature of X. Hence we shall write (as we did in the above example), $P(X = 1) = \frac{1}{2}$. What is meant is that a certain event in the sample space S, namely $\{HT, TH\} = \{s \mid X(s) = 1\}$ occurs with probability $\frac{1}{2}$. Hence we assign that same probability to the event $\{X = 1\}$ in the range space. We shall continue to write expressions like $P(X = 1)$, $P(X \leq 5)$, etc. It is *very important* for the reader to realize what these expressions really represent.

Once the probabilities associated with various outcomes (or events) in the range space R_X have been determined (more precisely, induced) we shall often ignore the original sample space S which gave rise to these probabilities. Thus in the above example, we shall simply be concerned with $R_X = \{0, 1, 2\}$ and the associated probabilities $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$. The fact that these probabilities are *determined* by a probability function defined over the original sample space S need not concern us if we are simply interested in studying the *values* of the random variable X.

In discussing, in detail, many of the important concepts associated with random variables, we shall find it convenient to distinguish between two important cases: the discrete and the continuous random variables.

4.2 Discrete Random Variables

Definition. Let X be a random variable. If the number of possible values of X (that is, R_X, the range space) is finite or countably infinite, we call X a *discrete random variable.* That is, the possible *values* of X may be listed as $x_1, x_2, \ldots, x_n, \ldots$ In the finite case the list terminates and in the countably infinite case the list continues indefinitely.

EXAMPLE 4.5. A radioactive source is emitting α-particles. The emission of these particles is observed on a counting device during a specified period of time. The following random variable is of interest:

$$X = \text{number of particles observed.}$$

What are the possible values of X? We shall assume that these values consist of all nonnegative integers. That is, $R_X = \{0, 1, 2, \ldots, n, \ldots\}$. An objection which we confronted once before may again be raised at this point. It could be argued that during a specified (finite) time interval it is impossible to observe more than, say N particles, where N may be a very large positive integer. Hence the possible values for X should really be: $0, 1, 2, \ldots, N$. However, it turns out to be mathematically simpler to consider the idealized description given above. In fact, whenever we assume that the possible values of a random variable X are countably infinite, we are actually considering an idealized representation of X.

In view of our previous discussions of the probabilistic description of events with a finite or countably infinite number of members, the probabilistic description of a discrete random variable will not cause any difficulty. We proceed as follows.

Definition. Let X be a discrete random variable. Hence R_X, the range space of X, consists of at most a countably infinite number of values, $x_1, x_2, \ldots$ With each possible outcome x_i we associate a number $p(x_i) = P(X = x_i)$, called the probability of x_i. The numbers $p(x_i)$, $i = 1, 2, \ldots$ must satisfy the following conditions:

$$\text{(a)} \quad p(x_i) \geq 0 \quad \text{for all } i,$$

$$\text{(b)} \quad \sum_{i=1}^{\infty} p(x_i) = 1. \tag{4.3}$$

The function p defined above is called the *probability function* (or point probability function) of the random variable X. The collection of pairs $(x_i, p(x_i))$, $i = 1, 2, \ldots$, is sometimes called the *probability distribution* of X.

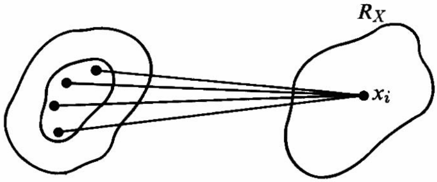

FIGURE 4.3

Notes: (a) The particular choice of the numbers $p(x_i)$ is presumably determined from the probability function associated with events in the sample space S on which X is defined. That is, $p(x_i) = P[s \mid X(s) = x_i]$. (See Eqs. 4.1 and 4.2.) However, since we are interested only in the values of X, that is R_X, and the probabilities associated with these values, we are again suppressing the functional nature of X. (See Fig. 4.3.) Although in most cases the numbers will in fact be determined from the probability distribution in some underlying sample space S, *any* set of numbers $p(x_i)$ satisfying Eq. (4.3) may serve as proper probabilistic description of a discrete random variable.

(b) If X assumes only a finite number of values, say $x_1, \ldots, x_N$, then $p(x_i) = 0$ for $i > N$, and hence the infinite series in Eq. (4.3) becomes a finite sum.

(c) We may again note an analogy to mechanics by considering a total mass of one unit distributed over the real line with the entire mass located at the points $x_1, x_2, \ldots$ The numbers $p(x_i)$ represent the amount of mass located at x_i.

(d) The geometric interpretation (Fig. 4.4) of a probability distribution is often useful.

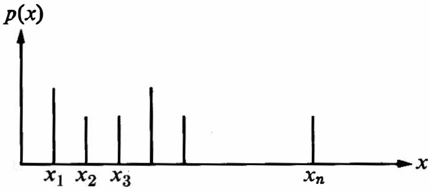

FIGURE 4.4

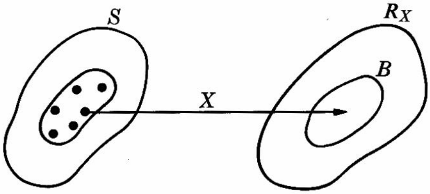

FIGURE 4.5

Let B be an event associated with the random variable X. That is, $B \subset R_X$ (Fig. 4.5). Specifically, suppose that $B = \{x_{i_1}, x_{i_2}, \ldots\}$. Hence

$$P(B) = P[s \mid X(s) \in B] \qquad \text{(since these events are equivalent)}$$

$$= P[s \mid X(s) = x_{i_j}, j = 1, 2, \ldots] = \sum_{j=1}^{\infty} p(x_{i_j}). \tag{4.4}$$

In words: The probability of an event B equals the sum of the probabilities of the individual outcomes associated with B.

Notes: (a) Suppose that the discrete random variable X may assume only a finite number of values, say $x_1, \ldots, x_N$. If each outcome is equally probable, then we obviously have $p(x_1) = \cdots = p(x_N) = 1/N$.

(b) If X assumes a countably infinite number of values, then it is *impossible* to have all outcomes equally probable. For we cannot possibly satisfy the condition $\sum_{i=1}^{\infty} p(x_i) = 1$ if we must have $p(x_i) = c$ for all i.

(c) In every finite interval there will be at most a finite number of the possible values of X. If some such interval contains none of these possible values we assign probability zero to it. That is, if $R_X = \{x_1, x_2, \ldots, x_n\}$ and if no $x_i \in [a, b]$, then $P[a \leq X \leq b] = 0$.

EXAMPLE 4.6. Suppose that a radio tube is inserted into a socket and tested. Assume that the probability that it tests positive equals $\frac{3}{4}$; hence the probability that it tests negative is $\frac{1}{4}$. Assume furthermore that we are testing a large supply of such tubes. The testing continues until the first positive tube appears. Define the random variable X as follows: X is the number of tests required to terminate the experiment. The sample space associated with this experiment is

$$S = \{+, -+, --+, ---+, \ldots\}.$$

To determine the probability distribution of X we reason as follows. The possible values of X are $1, 2, \ldots, n, \ldots$ (we are obviously dealing with the idealized sample space). And $X = n$ if and only if the first $(n - 1)$ tubes are negative and the nth tube is positive. If we suppose that the condition of one tube does not affect the condition of another, we may write

$$p(n) = P(X = n) = (\tfrac{1}{4})^{n-1}(\tfrac{3}{4}), \qquad n = 1, 2, \ldots$$

To check that these values of $p(n)$ satisfy Eq. (4.3) we note that

$$\sum_{n=1}^{\infty} p(n) = \frac{3}{4}\left(1 + \frac{1}{4} + \frac{1}{16} + \cdots\right)$$
$$= \frac{3}{4}\frac{1}{1 - \frac{1}{4}} = 1.$$

Note: We are using here the result that the *geometric series* $1 + r + r^2 + \cdots$ converges to $1/(1 - r)$ whenever $|r| < 1$. This is a result to which we shall refer repeatedly. Suppose that we want to evaluate $P(A)$, where A is defined as {The experiment ends after an even number of repetitions}. Using Eq. (4.4), we have

$$P(A) = \sum_{n=1}^{\infty} p(2n) = \frac{3}{16} + \frac{3}{256} + \cdots$$
$$= \frac{3}{16}(1 + \tfrac{1}{16} + \cdots)$$
$$= \frac{3}{16}\frac{1}{1 - \frac{1}{16}} = \frac{1}{5}.$$

4.3 The Binomial Distribution

In later chapters we shall consider, in considerable detail, a number of important discrete random variables. For the moment we shall simply study one of these and then use it to illustrate a number of important concepts.

EXAMPLE 4.7. Suppose that items coming off a production line are classified as defective (D) or nondefective (N). Suppose that three items are chosen at random from a day's production and are classified according to this scheme. The sample space for this experiment, say S, may be described as follows:

$$S = \{DDD, DDN, DND, NDD, NND, NDN, DNN, NNN\}.$$

(Another way of describing S is as $S = S_1 \times S_2 \times S_3$, the Cartesian product of S_1, S_2, and S_3, where each $S_i = \{D, N\}$.)

Let us suppose that with probability 0.2 an item is defective and hence with probability 0.8 an item is nondefective. Let us assume that these probabilities are the *same* for each item at least throughout the duration of our study. Finally let us suppose that the classification of any particular item is independent of the classification of any other item. Using these assumptions, it follows that the probabilities associated with the various outcomes of the sample space S as described above are

$$(0.2)^3, (0.8)(0.2)^2, (0.8)(0.2)^2, (0.8)(0.2)^2, (0.2)(0.8)^2, (0.2)(0.8)^2, (0.2)(0.8)^2, (0.8)^3.$$

Our interest usually is not focused on the individual outcomes of S. Rather, we simply wish to know *how many* defectives were found (irrespective of the order in which they occurred). That is, we wish to consider the random variable X which assigns to each outcome $s \in S$ the number of defectives found in s. Hence the set of possible values of X is $\{0, 1, 2, 3\}$.

We can obtain the probability distribution for X, $p(x_i) = P(X = x_i)$ as follows:

$$\begin{aligned}
X &= 0 \quad \text{if and only if} \quad NNN \text{ occurs;}\\
X &= 1 \quad \text{if and only if} \quad DNN,\ NDN, \text{ or } NND \text{ occurs;}\\
X &= 2 \quad \text{if and only if} \quad DDN,\ DND, \text{ or } NDD \text{ occurs;}\\
X &= 3 \quad \text{if and only if} \quad DDD \text{ occurs.}
\end{aligned}$$

(Note that $\{NNN\}$ is equivalent to $\{X = 0\}$, etc.) Hence

$$p(0) = P(X = 0) = (0.8)^3, \qquad p(1) = P(X = 1) = 3(0.2)(0.8)^2,$$
$$p(2) = P(X = 2) = 3(0.2)^2(0.8), \qquad p(3) = P(X = 3) = (0.2)^3.$$

Observe that the sum of these probabilities equals 1, for the sum may be written as $(0.8 + 0.2)^3$.

Note: The above discussion illustrates how the probabilities in the range space R_X (in this case $\{0, 1, 2, 3\}$) are *induced* by the probabilities defined over the sample space S. For the assumption that the eight outcomes of

$$S = \{DDD, DDN, DND, NDD, NND, NDN, DNN, NNN\}$$

have the probabilities given in Example 4.7, *determined* the value of $p(x)$ for all $x \in R_X$.

Let us now generalize the notions introduced in the above example.

Definition. Consider an experiment $\mathcal{E}$ and let A be some event associated with $\mathcal{E}$. Suppose that $P(A) = p$ and hence $P(\overline{A}) = 1 - p$. Consider n independent repetitions of $\mathcal{E}$. Hence the sample space consists of all possible sequences $\{a_1, a_2, \ldots, a_n\}$, where each a_i is either A or $\overline{A}$, depending on whether A or $\overline{A}$ occurred on the ith repetition of $\mathcal{E}$. (There are 2^n such sequences.) Furthermore, assume that $P(A) = p$ remains the same for all repetitions. Let the random variable X be defined as follows: X = number of times the event A occurred. We call X a *binomial* random variable with parameters n and p. Its possible values are obviously $0, 1, 2, \ldots, n$. (Equivalently we say that X has a *binomial distribution.*) The individual repetitions of $\mathcal{E}$ will be called *Bernoulli trials.*

Theorem 4.1. Let X be a binomial variable based on n repetitions. Then

$$P(X = k) = \binom{n}{k} p^k(1 - p)^{n-k}, \qquad k = 0, 1, \ldots, n. \tag{4.5}$$

Proof: Consider a particular element of the sample space of $\mathcal{E}$ satisfying the condition that $X = k$. One such outcome would arise, for instance, if the first k repetitions of $\mathcal{E}$ resulted in the occurrence of A, while the last $n - k$ repetitions resulted in the occurrence of $\overline{A}$, that is

$$\underbrace{AAA \cdots A}_{k}\ \underbrace{\overline{A}\overline{A}\overline{A} \cdots \overline{A}}_{n-k}.$$

Since all repetitions are independent, the probability of this particular sequence would be $p^k(1 - p)^{n-k}$. But exactly the same probability would be associated with any other outcome for which $X = k$. The total number of such outcomes equals $\binom{n}{k}$, for we must choose exactly k positions (out of n) for the A's. But this yields the above result, since these $\binom{n}{k}$ outcomes are all mutually exclusive.

Notes: (a) To verify our calculation we note that, using the binomial theorem, we have $\sum_{k=0}^{n} P(X = k) = \sum_{k=0}^{n} \binom{n}{k} p^k(1 - p)^{n-k} = [p + (1 - p)]^n = 1^n = 1$, as it should be. Since the probabilities $\binom{n}{k} p^k(1 - p)^{n-k}$ are obtained by expanding the binomial expression $[p + (1 - p)]^n$, we call this the binomial distribution.

(b) Whenever we perform independent repetitions of an experiment and are interested only in a dichotomy—defective *or* nondefective, hardness above *or* below a certain standard, noise level in a communication system above *or* below a preassigned threshold—we are potentially dealing with a sample space on which we may define a binomial random variable. So long as the conditions of experimentation stay sufficiently uniform so that the probability of some attribute, say A, stays constant, we may use the above model.

(c) If n is small, the individual terms of the binomial distribution are relatively easy to compute. However, if n is reasonably large, these computations become rather cumbersome. Fortunately, the binomial probabilities have been tabulated. There are many such tabulations. (See Appendix.)

EXAMPLE 4.8. Suppose that a radio tube inserted into a certain type of set has a probability of 0.2 of functioning more than 500 hours. If we test 20 tubes, what is the probability that exactly k of these function more than 500 hours, $k = 0, 1, 2, \ldots, 20$?

If X is the number of tubes functioning more than 500 hours, we shall assume that X has a binomial distribution. Thus $P(X = k) = \binom{20}{k}(0.2)^k(0.8)^{20-k}$.

The following values may be read from Table 4.1.

TABLE 4.1.

$P(X = 0) = 0.012$	$P(X = 4) = 0.218$	$P(X = 8) = 0.022$
$P(X = 1) = 0.058$	$P(X = 5) = 0.175$	$P(X = 9) = 0.007$
$P(X = 2) = 0.137$	$P(X = 6) = 0.109$	$P(X = 10) = 0.002$
$P(X = 3) = 0.205$	$P(X = 7) = 0.055$	$P(X = k) = 0^+$ for $k \geq 11$
(The remaining probabilities are less than 0.001.)		

If we plot this probability distribution, we obtain the graph shown in Fig. 4.6. The pattern which we observe here is quite general: The binomial probabilities increase monotonically until they reach a maximum value and then decrease monotonically. (See Problem 4.8.)

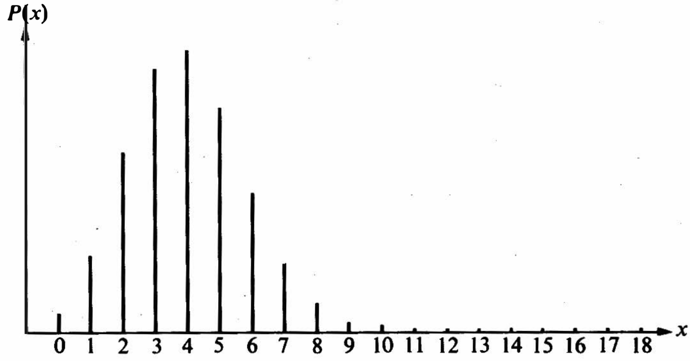

FIGURE 4.6

EXAMPLE 4.9. In operating a certain machine, there is a certain probability that the machine operator makes an error. It may be realistically assumed that the operator learns in the sense that the probability of his making an error decreases as he uses the machine repeatedly. Suppose that the operator makes n attempts and that the n trials are statistically independent. Suppose specifically that P(an error is made on the ith repetition) $= 1/(i+1)$, $i = 1, 2, \ldots, n$. Assume that 4 attempts are contemplated (that is, $n = 4$) and we define the random variable X as the number of machine operations made without error. Note that X is not binomially distributed because the probability of "success" is not constant.

To compute the probability that $X = 3$, for instance, we proceed as follows: $X = 3$ if and only if there is exactly one unsuccessful attempt. This can happen on the first, second, third, or fourth trial. Hence

$$P(X = 3) = \tfrac{1}{2}\tfrac{2}{3}\tfrac{3}{4}\tfrac{4}{5} + \tfrac{1}{2}\tfrac{1}{3}\tfrac{3}{4}\tfrac{4}{5} + \tfrac{1}{2}\tfrac{2}{3}\tfrac{1}{4}\tfrac{4}{5} + \tfrac{1}{2}\tfrac{2}{3}\tfrac{3}{4}\tfrac{1}{5} = \tfrac{5}{12}.$$

EXAMPLE 4.10. Consider a situation similar to the one described in Example 4.9. This time we shall assume that there is a constant probability p_1 of making no error on the machine during each of the first n_1 attempts and a constant probability $p_2 \le p_1$ of making no error on each of the next n_2 repetitions. Let X be the number of successful operations of the machine during the $n = n_1 + n_2$ independent attempts. Let us find a general expression for $P(X = k)$. For the same reason as given in the preceding example, X is not binomially distributed. To obtain $P(X = k)$ we proceed as follows.

Let Y_1 be the number of correct operations during the first n_1 attempts and let Y_2 be the number of correct operations during the second n_2 attempts. Hence Y_1 and Y_2 are independent random variables and $X = Y_1 + Y_2$. Thus $X = k$ if and only if $Y_1 = r$ and $Y_2 = k - r$, for any integer r satisfying $0 \le r \le n_1$ and $0 \le k - r \le n_2$.

The above restrictions on r are equivalent to $0 \le r \le n_1$ and $k - n_2 \le r \le k$. Combining these we may write

$$\max(0, k - n_2) \le r \le \min(k, n_1).$$

Hence we have

$$P(X = k) = \sum_{r=\max(0,k-n_2)}^{\min(k,n_1)} \binom{n_1}{r} p_1^r (1 - p_1)^{n_1 - r} \binom{n_2}{k - r} p_2^{k-r} (1 - p_2)^{n_2-(k-r)}.$$

With our usual convention that $\binom{a}{b} = 0$ whenever $b > a$ or $b < 0$, we may write the above probability as

$$P(X = k) = \sum_{r=0}^{n_1} \binom{n_1}{r} p_1^r (1 - p_1)^{n_1 - r} \binom{n_2}{k - r} p_2^{k-r} (1 - p_2)^{n_2-k+r}. \qquad (4.6)$$

For instance, if $p_1 = 0.2$, $p_2 = 0.1$, $n_1 = n_2 = 10$, and $k = 2$, the above probability becomes

$$P(X = 2) = \sum_{r=0}^{2} \binom{10}{r} (0.2)^r (0.8)^{10-r} \binom{10}{2-r} (0.1)^{2-r} (0.9)^{8+r} = 0.27,$$

after a straightforward calculation.

Note: Suppose that $p_1 = p_2$. In this case, Eq. (4.6) should reduce to $\binom{n}{k} p_1^k (1 - p_1)^{n-k}$, since now the random variable X *does* have a binomial distribution. To see that this is so, note that we may write (since $n_1 + n_2 = n$)

$$P(X = k) = p_1^k (1 - p_1)^{n-k} \sum_{r=0}^{n_1} \binom{n_1}{r} \binom{n_2}{k-r}.$$

To show that the above sum equals $\binom{n}{k}$ simply compare coefficients for the powers of x^k on both sides of the identity $(1 + x)^{n_1}(1 + x)^{n_2} = (1 + x)^{n_1+n_2}$.

4.4. Continuous Random Variables

Suppose that the range space of X is made up of a very large finite number of values, say all values x in the interval $0 \leq x \leq 1$ of the form 0, 0.01, 0.02, . . . , 0.98, 0.99, 1.00. With each of these values is associated a nonnegative number $p(x_i) = P(X = x_i)$, $i = 1, 2, \ldots$, whose sum equals 1. This situation is represented geometrically in Fig. 4.7.

We have pointed out before that it might be mathematically easier to idealize the above probabilistic description of X by supposing that X can assume *all* possible values, $0 \leq x \leq 1$. If we do this, what happens to the point probabilities $p(x_i)$? Since the possible values of X are noncountable, we cannot really speak of the ith value of X, and hence $p(x_i)$ becomes meaningless. What we shall do is to replace the function p, defined only for $x_1, x_2, \ldots$, by a function f defined (in the present context) for *all* values of x, $0 \leq x \leq 1$. The properties of Eq. (4.3) will be replaced by $f(x) \geq 0$ and $\int_0^1 f(x)\, dx = 1$. Let us proceed formally as follows.

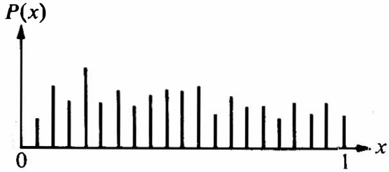

FIGURE 4.7

Definition. X is said to be a *continuous random variable* if there exists a function f, called the probability density function (pdf) of X, satisfying the following conditions:

(a) $f(x) \geq 0$ for all x,

(b) $$\int_{-\infty}^{+\infty} f(x)\, dx = 1. \tag{4.7}$$

(c) For any a, b, with $-\infty < a < b + \infty$, we have $P(a \leq X \leq b) = \int_a^b f(x)\,dx.$ (4.8)

Notes: (a) We are essentially saying that X is a continuous random variable if X may assume all values in some interval (c, d) where c and d may be $-\infty$ and $+\infty$, respectively. The stipulated existence of a pdf is a mathematical device which has considerable intuitive appeal and makes our computations simpler. In this connection it should again be pointed out that when we suppose that X is a continuous random variable, we are dealing with the *idealized* description of X.

(b) $P(c < X < d)$ represents the area under the graph in Fig. 4.8 of the pdf f between $x = c$ and $x = d$.

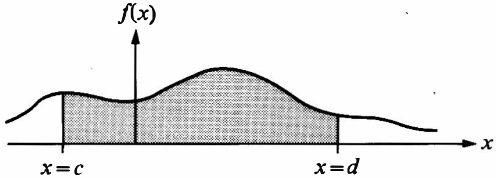

FIGURE 4.8

(c) It is a consequence of the above probabilistic description of X that for *any* specified value of X, say x_0, we have $P(X = x_0) = 0$, since $P(X = x_0) = \int_{x_0}^{x_0} f(x)\,dx = 0$. This result may seem quite contrary to our intuition. We must realize, however, that if we allow X to assume *all* values in some interval, then probability zero is not equivalent with impossibility. Hence in the continuous case, $P(A) = 0$ does not imply $A = \emptyset$, the empty set. (See Theorem 1.1.) Saying this more informally, consider choosing a point at random on the line segment $\{x \mid 0 \leq x \leq 2\}$. Although we might be willing to agree (for mathematical purposes) that *every* conceivable point on the segment could be the outcome of our experiment, we would be extremely surprised if in fact we chose precisely the midpoint of the segment, or any other *specified* point, for that matter. When we state this in precise mathematical language we say the event has "probability zero." In view of these remarks, the following probabilities are all the *same* if X is a continuous random variable:

$$P(c \leq X \leq d), \quad P(c \leq X < d), \quad P(c < X \leq d), \quad \text{and} \quad P(c < X < d).$$

(d) Although we shall not verify the details here, it may be shown that the above assignment of probabilities to events in R_X satisfies the basic axioms of probability (Eq. 1.3), where we may take $\{x \mid -\infty < x < +\infty\}$ as our sample space.

(e) If a function f^* satisfies the conditions, $f^*(x) \geq 0$, for all x, and $\int_{-\infty}^{+\infty} f^*(x)\,dx = K$, where K is a positive real number (not necessarily equal to 1), then f^* does *not* satisfy all the conditions for being a pdf. However, we may easily define a new function, say f, in terms of f^* as follows:

$$f(x) = \frac{f^*(x)}{K} \quad \text{for all } x.$$

Hence f satisfies all the conditions for a pdf.

(f) If X assumes values only in some finite interval $[a, b]$, we may simply set $f(x) = 0$ for all $x \notin [a, b]$. Hence the pdf is defined for *all* real values of x, and we may require that $\int_{-\infty}^{+\infty} f(x)\,dx = 1$. Whenever the pdf is specified only for certain values of x, we shall suppose that it is zero elsewhere.

(g) $f(x)$ does not represent the probability of anything! We have noted before that $P(X = 2) = 0$, for example, and hence $f(2)$ certainly does not represent this probability. Only when the function is integrated between two limits does it yield a probability. We can, however, give an interpretation of $f(x)\,\Delta x$ as follows. From the mean-value theorem of the calculus it follows that

$$P(x \leq X \leq x + \Delta x) = \int_x^{x+\Delta x} f(s)\,ds = \Delta x f(\xi), \qquad x \leq \xi \leq x + \Delta x.$$

If Δx is small, $f(x)\,\Delta x$ equals *approximately* $P(x \leq X \leq x + \Delta x)$. (If f is continuous from the right, this approximation becomes more accurate as $\Delta x \to 0$.)

(h) We should again point out that the probability distribution (in this case the pdf) is induced on R_X by the underlying probability associated with events in S. Thus, when we write $P(c < X < d)$, we mean, as always, $P[c < X(s) < d]$, which in turn equals $P[s \mid c < X(s) < d]$, since these events are equivalent. The above definition, Eq. (4.8), essentially stipulates the existence of a pdf f defined over R_X such that

$$P[s \mid c < X(s) < d] = \int_c^d f(x)\,dx.$$

We shall again suppress the functional nature of X and hence we shall be concerned only with R_X and the pdf f.

(i) In the continuous case we can again consider the following *analogy to mechanics:* Suppose that we have a total mass of one unit, continuously distributed over the interval $a \leq x \leq b$. Then $f(x)$ represents the mass density at the point x and $\int_c^d f(x)\,dx$ represents the total mass contained in the interval $c \leq x \leq d$.

EXAMPLE 4.11. The existence of a pdf was assumed in the above discussion of a continuous random variable. Let us consider a simple example in which we can easily determine the pdf by making an appropriate assumption about the probabilistic behavior of the random variable. Suppose that a point is chosen in the interval (0, 1). Let X represent the random variable whose value is the x-coordinate of the chosen point.

Assume: If I is any interval in (0, 1), then $\text{Prob}[X \in I]$ is directly proportional to the length of I, say $L(I)$. That is, $\text{Prob}[X \in I] = kL(I)$, where k is the constant of proportionality. (It is easy to see, by taking $I = (0, 1)$ and observing that $L\big((0, 1)\big) = 1$ and $\text{Prob}\,[X \in (0, 1)] = 1$, that $k = 1$.)

Obviously X assumes all values in (0, 1). What is its pdf? That is, can we find a function f such that

$$P(a < X < b) = \int_a^b f(x)\,dx?$$

Note that if $a < b < 0$ or $1 < a < b$, $P(a < X < b) = 0$ and hence $f(x) = 0$. If $0 < a < b < 1$, $P(a < X < b) = b - a$ and hence $f(x) = 1$. Thus we find,

$$f(x) = \begin{cases} 1, & 0 < x < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

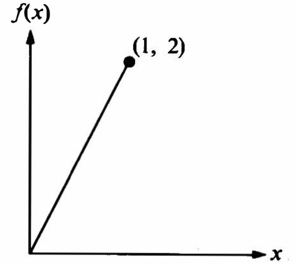

FIGURE 4.9

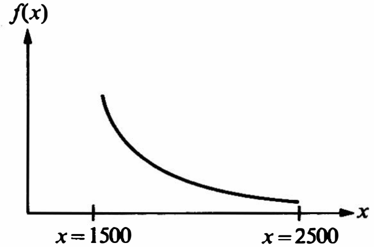

FIGURE 4.10

EXAMPLE 4.12. Suppose that the random variable X is continuous. (See Fig. 4.9.) Let the pdf f be given by

$$\begin{aligned} f(x) &= 2x, \quad 0 < x < 1, \\ &= 0, \quad \text{elsewhere.} \end{aligned}$$

Clearly, $f(x) \geq 0$ and $\int_{-\infty}^{+\infty} f(x)\,dx = \int_0^1 2x\,dx = 1$. To compute $P(X \leq \frac{1}{2})$, we must simply evaluate the integral $\int_0^{1/2} (2x)\,dx = \frac{1}{4}$.

The concept of conditional probability discussed in Chapter 3 can be meaningfully applied to random variables. For instance, in the above example we may evaluate $P(X \leq \frac{1}{2} \mid \frac{1}{3} \leq X \leq \frac{2}{3})$. Directly applying the definition of conditional probability, we have

$$\begin{aligned} P(X \leq \tfrac{1}{2} \mid \tfrac{1}{3} \leq X \leq \tfrac{2}{3}) &= \frac{P(\frac{1}{3} \leq X \leq \frac{1}{2})}{P(\frac{1}{3} \leq X \leq \frac{2}{3})} \\ &= \frac{\int_{1/3}^{1/2} 2x\,dx}{\int_{1/3}^{2/3} 2x\,dx} = \frac{5/36}{1/3} = \frac{5}{12}. \end{aligned}$$

EXAMPLE 4.13. Let X be the life length of a certain type of light bulb (in hours). Assuming X to be a continuous random variable, we suppose that the pdf f of X is given by

$$\begin{aligned} f(x) &= a/x^3, \quad 1500 \leq x \leq 2500, \\ &= 0, \quad \text{elsewhere.} \end{aligned}$$

(That is, we are assigning probability zero to the events $\{X < 1500\}$ and $\{X > 2500\}$.) To evaluate the constant a, we invoke the condition $\int_{-\infty}^{+\infty} f(x)\,dx =$

1 which in this case becomes $\int_{1500}^{2500} a/x^3\,dx = 1$. From this we obtain $a = 7,031,250$. The graph of f is shown in Fig. 4.10.

In a later chapter we shall study, in considerable detail, a number of important random variables, both discrete and continuous. We know from our use of deterministic models that certain functions play a far more important role than others. For example, the linear, quadratic, exponential, and trigonometric functions play a vital role in describing deterministic models. We shall find in developing nondeterministic (that is, probabilistic) models that certain random variables are of particular importance.

4.5 Cumulative Distribution Function

We want to introduce another important, general concept in this chapter.

Definition. Let X be a random variable, discrete or continuous. We define F to be the *cumulative distribution function* of the random variable X (abbreviated as cdf) where $F(x) = P(X \le x)$.

Theorem 4.2. (a) If X is a discrete random variable,

$$F(x) = \sum_j p(x_j), \tag{4.9}$$

where the sum is taken over all indices j satisfying $x_j \le x$.

(b) If X is a continuous random variable with pdf f,

$$F(x) = \int_{-\infty}^{x} f(s)\,ds. \tag{4.10}$$

Proof: Both of these results follow immediately from the definition.

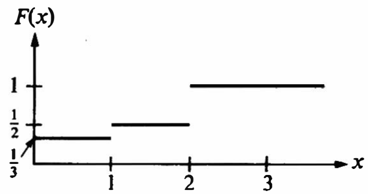

FIGURE 4.11

EXAMPLE 4.14. Suppose that the random variable X assumes the three values 0, 1, and 2 with probabilities $\frac{1}{3}$, $\frac{1}{6}$, and $\frac{1}{2}$, respectively. Then

$$\begin{aligned} F(x) &= 0 && \text{if } x < 0,\\ &= \tfrac{1}{3} && \text{if } 0 \le x < 1,\\ &= \tfrac{1}{2} && \text{if } 1 \le x < 2,\\ &= 1 && \text{if } x \ge 2. \end{aligned}$$

(Note that it is very important to indicate the inclusion or exclusion of the endpoints in describing the various intervals.) The graph of F is given in Fig. 4.11.

EXAMPLE 4.15. Suppose that X is a continuous random variable with pdf

$$\begin{aligned} f(x) &= 2x, \quad 0 < x < 1, \\ &= 0, \quad \text{elsewhere.} \end{aligned}$$

Hence the cdf F is given by

$$\begin{aligned} F(x) &= 0 \quad \text{if } x \leq 0, \\ &= \int_0^x 2s\, ds = x^2 \\ &\qquad \text{if } 0 < x \leq 1, \\ &= 1 \quad \text{if } x > 1. \end{aligned}$$

The graph is shown in Fig. 4.12.

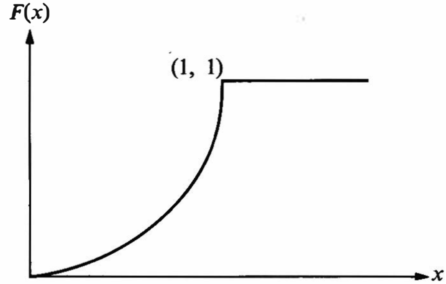

FIGURE 4.12

The graphs obtained in Figs. 4.11 and 4.12 for the cdf's are (in each case) quite typical in the following sense.

(a) If X is a discrete random variable with a finite number of possible values, the graph of the cdf F will be made up of horizontal line segments (it is called a step function). The function F is continuous except at the possible values of X, namely, $x_1, \ldots, x_n$. At the value x_j the graph will have a "jump" of magnitude $p(x_j) = P(X = x_j)$.

(b) If X is a continuous random variable, F will be a continuous function for all x.

(c) The cdf F is defined for *all* values of x, which is an important reason for considering it.

There are two important properties of the cdf which we shall summarize in the following theorem.

Theorem 4.3. (a) The function F is nondecreasing. That is, if $x_1 \leq x_2$, we have $F(x_1) \leq F(x_2)$.

(b) $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$. [We often write this as $F(-\infty) = 0, F(\infty) = 1$.]

Proof: (a) Define the events A and B as follows: $A = \{X \le x_1\}$, $B = \{X \le x_2\}$. Then, since $x_1 \le x_2$, we have $A \subset B$ and by Theorem 1.5, $P(A) \le P(B)$, which is the required result.

(b) In the continuous case we have

$$F(-\infty) = \lim_{x\to-\infty} \int_{-\infty}^{x} f(s)\,ds = 0,$$

$$F(\infty) = \lim_{x\to\infty} \int_{-\infty}^{x} f(s)\,ds = 1.$$

In the discrete case the argument is analogous.

The cumulative distribution function is important for a number of reasons. This is true particularly when we deal with a continuous random variable, for in that case we cannot study the probabilistic behavior of X by computing $P(X = x)$. That probability always equals zero in the continuous case. However, we can ask about $P(X \le x)$ and, as the next theorem demonstrates, obtain the pdf of X.

Theorem 4.4. (a) Let F be the cdf of a continuous random variable with pdf f. Then

$$f(x) = \frac{d}{dx} F(x),$$

for all x at which F is differentiable.

(b) Let X be a discrete random variable with possible values $x_1, x_2, \ldots,$ and suppose that it is possible to label these values so that $x_1 < x_2 < \cdots$ Let F be the cdf of X. Then

$$p(x_j) = P(X = x_j) = F(x_j) - F(x_{j-1}). \tag{4.12}$$

Proof: (a) $F(x) = P(X \le x) = \int_{-\infty}^{x} f(s)\,ds$. Thus applying the fundamental theorem of the calculus we obtain, $F'(x) = f(x)$.

(b) Since we assumed $x_1 < x_2 < \ldots,$ we have

$$F(x_j) = P(X = x_j \cup X = x_{j-1} \cup \cdots \cup X = x_1)$$
$$= p(j) + p(j-1) + \cdots + p(1).$$

And

$$F(x_{j-1}) = P(X = x_{j-1} \cup X = x_{j-2} \cup \cdots \cup X = x_1)$$
$$= p(j-1) + p(j-2) + \cdots + p(1).$$

Hence $F(x_j) - F(x_{j-1}) = P(X = x_j) = p(x_j)$.

Note: Let us briefly reconsider (a) of the above theorem. Recall the definition of the derivative of the function F:

$$F'(x) = \lim_{h\to 0} \frac{F(x+h) - F(x)}{h}$$

$$= \lim_{h\to 0^+} \frac{P(X \leq x+h) - P(X \leq x)}{h}$$

$$= \lim_{h\to 0^+} \frac{1}{h}[P(x < X \leq x+h)].$$

Thus if h is small and positive,

$$F'(x) = f(x) \cong \frac{P(x < X \leq x+h)}{h}.$$

That is, $f(x)$ is approximately equal to the "amount of probability in the interval $(x, x+h]$ per length h." Hence the name *probability density function.*

EXAMPLE 4.16. Suppose that a continuous random variable has cdf F given by

$$F(x) = 0, \quad x \leq 0,$$
$$= 1 - e^{-x}, \quad x > 0.$$

Then $F'(x) = e^{-x}$ for $x > 0$, and thus the pdf f is given by

$$f(x) = e^{-x}, \quad x \geq 0,$$
$$= 0, \quad \text{elsewhere.}$$

Note: A final word on terminology may be in order. This terminology, although not quite uniform, has become rather standardized. When we speak of the *probability distribution* of a random variable X we mean its pdf f if X is continuous, or its point probability function p defined for $x_1, x_2, \ldots$ if X is discrete. When we speak of the cumulative distribution function, or sometimes just the *distribution function,* we always mean F, where $F(x) = P(X \leq x)$.

4.6 Mixed Distributions

We have restricted our discussion entirely to random variables which are either discrete or continuous. Such random variables are certainly the most important in applications. However, there are situations in which we may encounter the *mixed* type: the random variable X may assume certain distinct values, say $x_1, \ldots, x_n$, with positive probability and also assume all values in some interval, say $a \leq x \leq b$. The probability distribution of such a random variable would be obtained by combining the ideas considered above for the description of discrete and continuous random variables as follows. To each value x_i assign a number $p(x_i)$ such that $p(x_i) \geq 0$, all i, and such that $\sum_{i=1}^{n} p(x_i) = p < 1$. Then define a function f satisfying $f(x) \geq 0$, $\int_a^b f(x)\,dx = 1 - p$. For all a, b, with $-\infty < a < b < +\infty$,

$$P(a \leq X \leq b) = \int_a^b f(x)\,dx + \sum_{\{i:a \leq x_i \leq b\}} p(x_i).$$

In this way we satisfy the condition

$$P(S) = P(-\infty < X < \infty) = 1.$$

A random variable of mixed type might arise as follows. Suppose that we are testing some equipment and we let X be the time of functioning. In most problems we would describe X as a continuous random variable with possible values $x \geq 0$. However, situations may arise in which there is a positive probability that the item does not function at all, that is, it fails at time $X = 0$. In such a case we would want to modify our model and assign a positive probability, say p, to the outcome $X = 0$. Hence we would have $P(X = 0) = p$ and $P(X > 0) = 1 - p$. Thus the number p would describe the distribution of X at 0, while the pdf f would describe the distribution for values of $X > 0$ (Fig. 4.13).

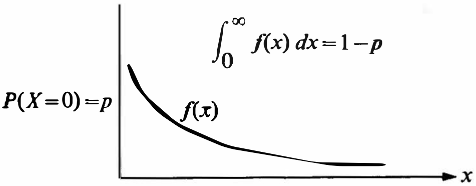

FIGURE 4.13

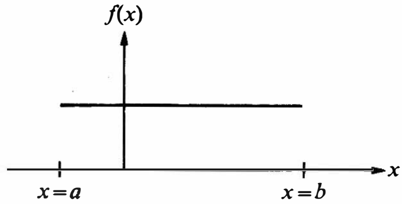

FIGURE 4.14

4.7 Uniformly Distributed Random Variables

In Chapters 8 and 9 we shall study in considerable detail a number of important discrete and continuous random variables. We have already introduced the important binomial random variable. Let us now consider briefly an important continuous random variable.

Definition. Suppose that X is a continuous random variable assuming all values in the interval $[a, b]$, where both a and b are finite. If the pdf of X is given by

$$f(x) = \frac{1}{b-a}, \quad a \leq x \leq b, \qquad (4.13)$$
$$= 0, \quad \text{elsewhere},$$

we say that X is *uniformly distributed* over the interval $[a, b]$. (See Fig. 4.14.)

Notes: (a) A uniformly distributed random variable has a pdf which is *constant* over the interval of definition. In order to satisfy the condition $\int_{-\infty}^{+\infty} f(x)\,dx = 1$, this constant must be equal to the reciprocal of the length of the interval.

(b) A uniformly distributed random variable represents the continuous analog to equally likely outcomes in the following sense. For any subinterval $[c, d]$, where $a \leq c < d \leq b$, $P(c \leq X \leq d)$ is the *same* for all subintervals having the same length. That is,

$$P(c \leq X \leq d) = \int_c^d f(x)\,dx = \frac{d-c}{b-a}$$

and thus depends only on the length of the interval and not on the location of that interval.

(c) We can now make precise the intuitive notion of *choosing a point P at random* on an interval, say $[a, b]$. By this we shall simply mean that the x-coordinate of the chosen point, say X, is uniformly distributed over $[a, b]$.

EXAMPLE 4.17. A point is chosen at random on the line segment $[0, 2]$. What is the probability that the chosen point lies between 1 and $\frac{3}{2}$?

Letting X represent the coordinate of the chosen point, we have that the pdf of X is given by $f(x) = \frac{1}{2}$, $0 < x < 2$, and hence $P(1 \leq X \leq \frac{3}{2}) = \frac{1}{4}$.

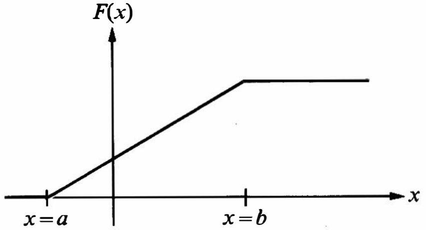

FIGURE 4.15

EXAMPLE 4.18. The hardness, say H, of a specimen of steel (measured on the Rockwell scale) may be assumed to be a continuous random variable uniformly distributed over [50, 70] on the B scale. Hence

$$\begin{aligned} f(h) &= \tfrac{1}{20}, && 50 < h < 70, \\ &= 0, && \text{elsewhere.} \end{aligned}$$

EXAMPLE 4.19. Let us obtain an expression for the cdf of a uniformly distributed random variable.

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(s)\,ds$$

$$\begin{aligned} &= 0 && \text{if } x < a, \\ &= \frac{x-a}{b-a} && \text{if } a \leq x < b, \\ &= 1 && \text{if } x \geq b. \end{aligned}$$

The graph is shown in Fig. 4.15.

4.8 A Remark

We have pointed out repeatedly that at some stage in our development of a probabilistic model, some probabilities must be assigned to outcomes on the basis of either experimental evidence (such as relative frequencies, for example) or some other considerations, such as past experience with the phenomena being studied. The following question might occur to the student: Why could we not obtain all probabilities we are interested in by some such nondeductive means? The answer is that many events whose probabilities we wish to know are so involved that our intuitive knowledge is insufficient. For instance, suppose that 1000 items are coming off a production line every day, some of which are defective. We wish to know the probability of having 50 or fewer defective items on a given day. Even if we are familiar with the general behavior of the production process, it might be difficult for us to associate a quantitative measure with the event: 50 or fewer items are defective. However, we might be able to make the statement that any individual item had probability 0.10 of being defective. (That is, past experience gives us the information that about 10 percent of the items are defective.) Furthermore, we might be willing to assume that individual items are defective or nondefective independently of one another. *Now* we can proceed deductively and *derive* the probability of the event under consideration. That is, if X = number of defectives,

$$P(X \leq 50) = \sum_{k=0}^{50} \binom{1000}{k} (0.10)^k (0.90)^{1000-k}.$$

The point being made here is that the various methods for computing probabilities which we have derived (and others which we shall study subsequently) are of great importance since with them we can evaluate probabilities associated with rather involved events which would be difficult to obtain by intuitive or empirical means.

PROBLEMS

4.1. A coin is known to come up heads three times as often as tails. This coin is tossed three times. Let X be the number of heads that appear. Write out the probability distribution of X and also the cdf. Make a sketch of both.

4.2. From a lot containing 25 items, 5 of which are defective, 4 are chosen at random. Let X be the number of defectives found. Obtain the probability distribution of X if

(a) the items are chosen with replacement,
(b) the items are chosen without replacement.

4.3. Suppose that the random variable X has possible values $1, 2, 3, \ldots,$ and $P(X = j) = 1/2^j, j = 1, 2, \ldots$

(a) Compute $P(X \text{ is even})$.
(b) Compute $P(X \geq 5)$.
(c) Compute $P(X \text{ is divisible by } 3)$.

4.4. Consider a random variable X with possible outcomes: 0, 1, 2, ... Suppose that $P(X = j) = (1 - a)a^j, j = 0, 1, 2, \ldots$
(a) For what values of a is the above model meaningful?
(b) Verify that the above does represent a legitimate probability distribution.
(c) Show that for any two positive integers s and t,

$$P(X > s + t \mid X > s) = P(X \geq t).$$

4.5. Suppose that twice as many items are produced (per day) by machine 1 as by machine 2. However, about 4 percent of the items from machine 1 tend to be defective while machine 2 produces only about 2 percent defectives. Suppose that the daily output of the two machines is combined. A random sample of 10 is taken from the combined output. What is the probability that this sample contains 2 defectives?

4.6. Rockets are launched until the first successful launching has taken place. If this does not occur within 5 attempts, the experiment is halted and the equipment inspected. Suppose that there is a constant probability of 0.8 of having a successful launching and that successive attempts are independent. Assume that the cost of the first launching is K dollars while subsequent launchings cost $K/3$ dollars. Whenever a successful launching takes place, a certain amount of information is obtained which may be expressed as financial gain of, say C dollars. If T is the net cost of this experiment, find the probability distribution of T.

4.7. Evaluate $P(X = 5)$, where X is the random variable defined in Example 4.10. Suppose that $n_1 = 10$, $n_2 = 15$, $p_1 = 0.3$ and $p_2 = 0.2$.

4.8. (*Properties of the binomial probabilities.*) In the discussion of Example 4.8 a general pattern for the binomial probabilities $\binom{n}{k} p^k (1 - p)^{n-k}$ was suggested. Let us denote these probabilities by $p_n(k)$.
(a) Show that for $0 \leq k < n$ we have

$$p_n(k + 1)/p_n(k) = [(n - k)/(k + 1)]\,[p/(1 - p)].$$

(b) Using (a) show that

$$\begin{aligned}
&\text{(i) } p_n(k + 1) > p_n(k) \quad &&\text{if} \quad k < np - (1 - p),\\
&\text{(ii) } p_n(k + 1) = p_n(k) \quad &&\text{if} \quad k = np - (1 - p),\\
&\text{(iii) } p_n(k + 1) < p_n(k) \quad &&\text{if} \quad k > np - (1 - p).
\end{aligned}$$

(c) Show that if $np - (1 - p)$ is an integer, $p_n(k)$ assumes its maximum value for two values of k, namely $k_0 = np - (1 - p)$ and $k_0' = np - (1 - p) + 1$.
(d) Show that if $np - (1 - p)$ is not an integer then $p_n(k)$ assumes its maximum value when k is equal to the smallest integer greater than k_0.
(e) Show that if $np - (1 - p) < 0$, $p_n(0) > p_n(1) > \cdots > p_n(n)$ while if $np - (1 - p) = 0$, $p_n(0) = p_n(1) > p_n(2) > \cdots > p_n(n)$.

4.9. The continuous random variable X has pdf $f(x) = x/2$, $0 \leq x \leq 2$. Two independent determinations of X are made. What is the probability that both these determinations will be greater than one? If three independent determinations had been made, what is the probability that exactly two of these are larger than one?

4.10. Let X be the life length of an electron tube and suppose that X may be represented as a continuous random variable with pdf $f(x) = be^{-bx}$, $x \geq 0$. Let $p_j = P(j \leq X < j + 1)$. Show that p_j is of the form $(1 - a)a^j$ and determine a.

4.11. The continuous random variable X has pdf $f(x) = 3x^2$, $-1 \leq x \leq 0$. If b is a number satisfying $-1 < b < 0$, compute $P(X > b \mid X < b/2)$.

4.12. Suppose that f and g are pdf's on the same interval, say $a \leq x \leq b$.

(a) Show that $f + g$ is *not* a pdf on that interval.

(b) Show that for every number β, $0 < \beta < 1$, $\beta f(x) + (1 - \beta)g(x)$ is a pdf on that interval.

4.13. Suppose that the graph in Fig. 4.16 represents the pdf of a random variable X.

(a) What is the relationship between a and b?

(b) If $a > 0$ and $b > 0$, what can you say about the largest value which b may assume? (See Fig. 4.16.)

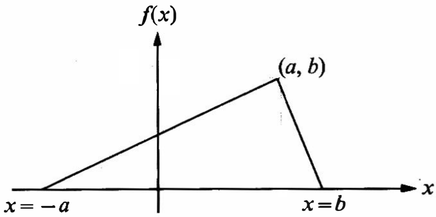

FIGURE 4.16

4.14. The percentage of alcohol $(100X)$ in a certain compound may be considered as a random variable, where X, $0 < X < 1$, has the following pdf:

$$f(x) = 20x^3(1 - x), \qquad 0 < x < 1.$$

(a) Obtain an expression for the cdf F and sketch its graph.

(b) Evaluate $P(X \leq \frac{2}{3})$.

(c) Suppose that the selling price of the above compound depends on the alcohol content. Specifically, if $\frac{1}{3} < X < \frac{2}{3}$, the compound sells for C_1 dollars/gallon; otherwise it sells for C_2 dollars/gallon. If the cost is C_3 dollars/gallon, find the probability distribution of the net profit per gallon.

4.15. Let X be a continuous random variable with pdf f given by:

$$\begin{aligned} f(x) &= ax, & 0 \leq x \leq 1, \\ &= a, & 1 \leq x \leq 2, \\ &= -ax + 3a, & 2 \leq x \leq 3, \\ &= 0, & \text{elsewhere}. \end{aligned}$$

(a) Determine the constant a. (b) Determine F, the cdf, and sketch its graph.

(c) If X_1, X_2, and X_3 are three independent observations from X, what is the probability that exactly one of these three numbers is larger than 1.5?

4.16. The diameter on an electric cable, say X, is assumed to be a continuous random variable with pdf $f(x) = 6x(1 - x)$, $0 \leq x \leq 1$.

(a) Check that the above is a pdf and sketch it.
(b) Obtain an expression for the cdf of X and sketch it.
(c) Determine a number b such that $P(X < b) = 2P(X > b)$.
(d) Compute $P(X \le \frac{1}{2} \mid \frac{1}{3} < X < \frac{2}{3})$.

4.17. Each of the following functions represents the cdf of a continuous random variable. In each case $F(x) = 0$ for $x < a$ and $F(x) = 1$ for $x > b$, where $[a, b]$ is the indicated interval. In each case, sketch the function F, determine the pdf f and sketch it. Also verify that f is a pdf.

(a) $F(x) = x/5, 0 \le x \le 5$ (b) $F(x) = (2/\pi) \sin^{-1}(\sqrt{x}), 0 \le x \le 1$
(c) $F(x) = e^{3x}, -\infty < x \le 0$ (d) $F(x) = x^3/2 + \frac{1}{2}, -1 \le x \le 1.$

4.18. Let X be the life length of an electronic device (measured in hours). Suppose that X is a continuous random variable with pdf $f(x) = k/x^n$, $2000 \le x \le 10{,}000$.

(a) For $n = 2$, determine k.
(b) For $n = 3$, determine k.
(c) For general n, determine k.
(d) What is the probability that the device will fail before 5000 hours have elapsed?
(e) Sketch the cdf $F(t)$ for (c) and determine its algebraic form.

4.19. Let X be a binomially distributed random variable based on 10 repetitions of an experiment. If $p = 0.3$, evaluate the following probabilities using the table of the binomial distribution in the Appendix.

(a) $P(X \le 8)$ (b) $P(X = 7)$ (c) $P(X > 6)$.

4.20. Suppose that X is uniformly distributed over $[-\alpha, +\alpha]$, where $\alpha > 0$. Whenever possible, determine α so that the following are satisfied.

(a) $P(X > 1) = \frac{1}{3}$ (b) $P(X > 1) = \frac{1}{2}$ (c) $P(X < \frac{1}{2}) = 0.7$
(d) $P(X < \frac{1}{2}) = 0.3$ (e) $P(|X| < 1) = P(|X| > 1)$.

4.21. Suppose that X is uniformly distributed over $[0, \alpha]$, $\alpha > 0$. Answer the questions of Problem 4.20.

4.22. A point is chosen at random on a line of length L. What is the probability that the ratio of the shorter to the longer segment is less than $\frac{1}{4}$?

4.23. A factory produces 10 glass containers daily. It may be assumed that there is a constant probability $p = 0.1$ of producing a defective container. Before these containers are stored they are inspected and the defective ones are set aside. Suppose that there is a constant probability $r = 0.1$ that a defective container is misclassified. Let X equal the number of containers classified as defective at the end of a production day. (Suppose that all containers which are manufactured on a particular day are also inspected on that day.)

(a) Compute $P(X = 3)$ and $P(X > 3)$. (b) Obtain an expression for $P(X = k)$.

4.24. Suppose that 5 percent of all items coming off a production line are defective. If 10 such items are chosen and inspected, what is the probability that at most 2 defectives are found?

4.25. Suppose that the life length (in hours) of a certain radio tube is a continuous random variable X with pdf $f(x) = 100/x^2$, $x > 100$, and 0 elsewhere.

(a) What is the probability that a tube will last less than 200 hours if it is known that the tube is still functioning after 150 hours of service?

(b) What is the probability that if 3 such tubes are installed in a set, exactly one will have to be replaced after 150 hours of service?

(c) What is the maximum number of tubes that may be inserted into a set so that there is a probability of 0.5 that after 150 hours of service all of them are still functioning?

4.26. An experiment consists of n independent trials. It may be supposed that because of "learning," the probability of obtaining a successful outcome increases with the number of trials performed. Specifically, suppose that P(success on the ith repetition) $= (i+1)/(i+2)$, $i = 1, 2, \ldots, n$.

(a) What is the probability of having at least 3 successful outcomes in 8 repetitions?

(b) What is the probability that the first successful outcome occurs on the eighth repetition?

4.27. Referring to Example 4.10,

(a) evaluate $P(X = 2)$ if $n = 4$,

(b) for arbitrary n, show that $P(X = n - 1) = P$ (exactly one unsuccessful attempt) is equal to $[1/(n+1)] \sum_{i=1}^{n} (1/i)$.

4.28. If the random variable K is uniformly distributed over $(0, 5)$, what is the probability that the roots of the equation $4x^2 + 4xK + K + 2 = 0$ are real?

4.29. Suppose that the random variable X has possible values $1, 2, 3, \ldots$ and that $P(X = r) = k(1 - \beta)^{r-1}$, $0 < \beta < 1$.

(a) Determine the constant k.

(b) Find the *mode* of this distribution (i.e., that value of r which makes $P(X = r)$ largest).

4.30. A random variable X may assume four values with probabilities $(1 + 3x)/4$, $(1 - x)/4$, $(1 + 2x)/4$, and $(1 - 4x)/4$. For what values of x is this a probability distribution?