Exercises 3 through 8 are devoted to providing proofs of the four properties (a), (b), (c), (d) of countable sets listed in Section 13.19.

3. Prove that every subset of a countable set is countable. [*Hint:* Suppose S is a countably infinite set, say $S = \{x_1, x_2, x_3, \ldots\}$, and let A be an infinite subset of S. Let k(1) be the smallest positive integer m such that $x_m \in A$. Assuming k(l), k(2), . . . , $k(n - 1)$ have been defined, let k(n) be the smallest positive integer $m > k(n - 1)$ such that $x_m \in A$. Let $f(n) = x_{k(n)}$. Show that f is a one-to-one function whose domain is the set of positive integers and whose range is A. This proves the result when S is countably infinite. Construct a separate proof for a finite S.]
4. Show that the intersection of any collection of countable sets is countable. [Hint: Use the result of Exercise 3.]
5. Let $P = \{1, 2, 3, \ldots\}$ denote the set of positive integers.

   (a) Prove that the Cartesian product $P \times P$ is countable. [Hint: Let Q denote the set of positive integers of the form 2^m3^n, where m and n are positive integers. Then $Q \subset P$, so Q is countable (by Exercise 3). If $(m, n) \in P \times P$, let $f(m, n) = 2^m3^n$ and use this function to show that $P \times P \sim Q$.]

   (b) Deduce from part (a) that the Cartesian product of two countable sets is countable. Then use induction to extend the result to n countable sets.
6. Let $\mathscr{B} = \{B_1, B_2, B_3, \ldots\}$ be a countable collection of *disjoint* sets ($B_i \cap B_j = \varnothing$ when $i \neq j$) such that each B_n is countable. Show that the union $\bigcup_{k=1}^{\infty} B_k$ is also countable. [*Hint:* Let $B_n = \{b_{1,n}, b_{2,n}, b_{3,n}, \ldots\}$ and $S = \bigcup_{k=1}^{\infty} B_k$. If $x \in S$, then $x = b_{m,n}$ for some unique pair (m, n) and we can define $f(x) = (m, n)$. Use this f to show that S is equivalent to a subset of $P \times P$ and deduce (by Exercise 5) that S is countable.]
7. Let $\mathscr{A} = \{A_1, A_2, A_3, \ldots\}$ be a countable collection of sets, and let $\mathscr{B} = \{B_1, B_2, B_3, \ldots\}$ be defined as follows: $B_1 = A_1$, and, for $n > 1$,

$$B_n = A_n - \bigcup_{k=1}^{n-1} A_k .$$

   That is, B_n consists of those points in A_n which are not in any of the earlier sets $A_1, \ldots, A_{n-1}$. Prove that $\mathscr{B}$ is a collection of disjoint sets ($B_i \cap B_j = \varnothing$ when $i \neq j$) and that

$$\bigcup_{k=1}^{\infty} A_k = \bigcup_{k=1}^{\infty} B_k .$$

   This enables us to express the union of any countable collection of sets as the union of a countable collection of *disjoint* sets.
8. If $\mathscr{F}$ is a countable collection of countable sets, prove that the union of all sets in $\mathscr{F}$ is countable. [*Hint:* Use Exercises 6 and 7.]
9. Show that the following sets are countable:

   (a) The set of all intervals on the real axis with rational end points.

   (b) The set of all circles in the plane with rational radii and centers having rational coordinates.

   (c) Any set of disjoint intervals of positive length.
10. Show that the following sets are uncountable:

    (a) The set of irrational numbers in the interval (0, 1).

    (b) The set of all intervals of positive length.

    (c) The set of all sequences whose terms are the integers 0 and 1. (Recall that a sequence is a function whose domain is the set of positive integers.)

13.21 The definition of probability for countably infinite sample spaces

This section extends the definition of probability to countably infinite sample spaces. Let S be a countably infinite set and let $\mathscr{B}$ be a Boolean algebra of subsets of S. We define a probability measure P on $\mathscr{B}$ as we did for the finite case, except that we require countable additivity as well as finite additivity. That is, for every countably infinite collection $\{A_1, A_2, \ldots\}$ of elements of $\mathscr{B}$, we require that

$$(13.25) \qquad P\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k) \qquad \text{if } A_i \cap \text{A,} = \mathbf{0} \qquad \text{whenever } i \neq j.$$

Finitely additive set functions which satisfy (13.25) are said to be *countably additive* (or completely additive). Of course, this property requires assuming also that the countable union A, ∪ A, ∪ A, ∪ . · · is in $\mathscr{B}$ whenever each A, is in $\mathscr{B}$. Not all Boolean algebras have this property. Those which do are called Boolean c-algebras. An example is the Boolean algebra of all subsets of S.

DEFINITION OF PROBABILITY FOR COUNTABLY INFINITE SAMPLE SPACES. *Let $\mathscr{B}$ denote a Boolean o-algebra whose elements are subsets of a given countably infinite set S. A set function P is called a probability measure on $\mathscr{B}$ if it is nonnegative, countably additive, and satisfies P(S) = 1 .*

When $\mathscr{B}$ is the Boolean algebra of all subsets of S, a probability function is completely determined by its values on the singletons (called point probabilities). Every subset A of S is either finite or countably infinite, and the probability of A is computed by adding the point probabilities for all elements in A,

$$P(A) = \sum_{x \in A} P(x).$$

The sum on the right is either a finite sum or an absolutely convergent infinite series.

The following example illustrates an experiment with a countably infinite sample space.

EXAMPLE. Toss a coin repeatedly until the first outcome occurs a second time; at this point the game ends.

For a sample space we take the collection of all possible games that can be played. This set can be expressed as the union of two countably infinite sets A and B, where

$$A = \{TT, THT, THHT, THHHT, \ldots\} \quad \text{and} \quad B = (HH, HTH, HTTH, HTTTH, \ldots\}.$$

We denote the elements of set A (in the order listed) as a,, a,, a,, . . ., and those of set B as b,, $b_1, b_2, \ldots$. We can assign arbitrary nonnegative point probabilities $P(a_n)$ and $P(b_n)$ provided that

$$\sum_{n=0}^{\infty} P(a_n) + \sum_{n=0}^{\infty} P(b_n) = 1.$$

For example, suppose the coin has probability p of coming up heads *(H)* and probability $q = 1 - p$ of coming up tails (T), where $0 < p < 1$. Then a natural assignment of point probabilities would be

(13.26) $$P(a_n) = q^2p^n \quad \text{and} \quad P(b_n) = p^2q^n.$$

This is an acceptable assignment of probabilities because we have

$$\sum_{n=0}^{\infty} P(a_n) + \sum_{n=0}^{\infty} P(b_n) = q^2\sum_{n=0}^{\infty} p^n + p^2\sum_{n=0}^{\infty} q^n = \frac{q^2}{1-p} + \frac{p^2}{1-q} = q + p = 1.$$

Now suppose we ask for the probability that the game ends after exactly $n + 2$ tosses. This is the event $\{a_n\} \cup \{b_n\}$, and its probability is

$$P(a_n) + P(b_n) = q^2p^n + p^2q^n.$$

The probability that the game ends in at most $n + 2$ tosses is

$$\sum_{k=0}^{n} P(a_k) + \sum_{k=0}^{n} P(b_k) = q^2\frac{1-p^{n+1}}{1-p} + p^2\frac{1-q^{n+1}}{1-q} = 1 - qp^{n+1} - pq^{n+1}.$$

13.22 Exercises

The exercises in this section refer to the example in Section 13.21.

1. Using the point probabilities assigned in Equation (13.26), let $f_n(p)$ denote the probability that the game ends after exactly $n + 2$ tosses. Calculate the absolute maximum and minimum values of $f_n(p)$ on the interval $0 \le p \le 1$ for each of the values $n = 0, 1, 2, 3$.
2. Show that each of the following is an acceptable assignment of point probabilities.

   (a) $P(a_n) = P(b_n) = \dfrac{1}{2^{n+2}}$ for $n = 0, 1, 2, \ldots.$

   (b) $P(a_n) = P(b_n) = \dfrac{1}{(n+2)(n+3)}$ for $n = 0, 1, 2, \ldots.$

3. Calculate the probability that the game ends before the fifth toss, using:
   (a) the point probabilities in (13.26).
   (b) the point probabilities in Exercise 2(a).
   (c) the point probabilities in Exercise 2(b).
4. Calculate the probability that an odd number of tosses is required to terminate the game, using:
   (a) the point probabilities in (13.26).
   (b) the point probabilities in Exercise 2(a).
   (c) the point probabilities in Exercise 2(b).

13.23 Miscellaneous exercises on probability

1. What is the probability of rolling a ten with two unbiased dice?
2. Ten men and their wives are seated at random at a banquet. Compute the probability that a particular man sits next to his wife if (a) they are seated at a round table; (b) they are seated in a row.

3. A box has two drawers. Drawer number 1 contains four gold coins and two silver coins. Drawer number 2 contains three gold coins and three silver coins. A drawer is opened at random and a coin selected at random from the open drawer. Compute the probability of each of the following events:
(a) Drawer number 2 was opened and a silver coin was selected.
(b) A gold coin was selected from the opened drawer.
4. Two cards are picked in succession from a deck of 52 cards, each card having the same probability of being drawn.
(a) What is the probability that at least one is a spade?
The two cards are placed in a sack unexamined. One card is drawn from the sack and examined and found not to be a spade. (Each card has the same probability of being drawn.)
(b) What is the probability now of having at least one spade?
The card previously drawn is replaced in the sack and the cards mixed. Again a card is drawn and examined. No comparison is made to see if it is the same card previously drawn. The card is again replaced in the sack and the cards mixed. This is done a total of three times, including that of part (b), and each time the card examined is not a spade.
(c) What is a sample space and a probability function for this experiment? What is the probability that one of the two original cards is a spade?
5. A man has ten pennies, 9 ordinary and 1 with two heads. He selects a penny at random, tosses it six times, and it always comes up heads. Compute the probability that he selected the double-headed penny.
6. Prove that it is impossible to load a pair of dice so that every outcome from 2 to 12 will have the same probability of occurrence.
7. A certain Caltech sophomore has an alarm clock which will ring at the appointed hour with probability 0.7. If it rings, it will wake him in time to attend his mathematics class with probability 0.8. If it doesn't ring he will wake in time to attend his class with probability 0.3. Compute the probability that he will wake in time to attend his mathematics class.
8. Three horses A, B, and C are in a horse race. The event "A beats B" will be denoted symbolically by writing AB. The event "A beats B who beats C" will be denoted by ABC, etc. Suppose it is known that
$$P(AB) = \tfrac{2}{3}, \qquad P(AC) = \tfrac{2}{3}, \qquad P(BC) = \tfrac{1}{2},$$
and that
$$P(ABC) = P(ACB), \qquad P(BAC) = P(BCA), \qquad P(CAB) = P(CBA).$$
(a) Compute the probability that A wins.
(b) Compute the probability that B wins.
(c) Compute the probability that C wins.
(d) Are the events AB, AC, and CB independent?
9. The final step in a long computation requires the addition of three integers a_1, a_2, a_3. Assume that (a) the computations of a_1, a_2, and a_3 are stochastically independent; (b) in the computation of each a_i there is a common probability p that it is correct and that the probability of making an error of $+1$ is equal to the probability of making an error of -1; (c) no error larger than $+1$ or less than -1 can occur. Remember the possibility of compensating errors, and compute the probability that the sum $a_1 + a_2 + a_3$ is correct.
10. *The game of "odd man out."* Suppose n persons toss identical coins simultaneously and independently, where $n \geq 3$. Assume there is a probability p of obtaining heads with each coin. Compute the probability that in a given toss there will be an "odd man," that is, a person whose coin does not have the same outcome as that of any other member of the group.
11. Suppose n persons play the game "odd man out" with fair coins (as described in Exercise 10). For a given integer m compute the probability that it will take exactly m plays to conclude the game (the mth play is the first time there is an "odd man").

12. Suppose a compound experiment $(S, \mathscr{B}, P)$ is determined by two stochastically independent experiments $(S_1, \mathscr{B}_1, P_1)$ and $(S_2, \mathscr{B}_2, P_2)$, where $S = S_1 \times S_2$ and

$$P(x, y) = P_1(x)P_2(y)$$

for each (x, y) in S. The purpose of this exercise is to establish the formula

$$P(U \times V) = P_1(U)P_2(V) \tag{13.27}$$

for every pair of subsets U in $\mathscr{B}_1$ and V in $\mathscr{B}_2$. The sample spaces S_1 and S_2 are assumed to be finite.

(a) Verify that Equation (13.27) is true when U and V are singletons, and also when at least one of U or V is empty.

Suppose now that

$$U = \{u_1, u_2, \ldots, u_k\} \quad \text{and} \quad V = \{v_1, v_2, \ldots, v_m\}.$$

Then $U \times V$ consists of the km pairs (u_i, v_j). For each $i = 1, 2, \ldots, k$ let A_i denote the set of m pairs in $U \times V$ whose first component is u_i.

(b) Show that the A_i are disjoint sets whose union is $U \times V$.

(c) Show that

$$P(A_i) = \sum_{j=1}^{m} P(u_i, v_j) = P_1(u_i)P_2(V).$$

(d) From (b) and (c) deduce that

$$P(U \times V) = \sum_{i=1}^{k} P(A_i) = P_1(U)P_2(V).$$

14

CALCULUS OF PROBABILITIES

14.1 The definition of probability for uncountable sample spaces

A line segment is broken into two pieces, with the point of subdivision chosen at random. What is the probability that the two pieces have equal length? What is the probability that the longer segment has exactly twice the length of the shorter? What is the probability that the longer segment has at least twice the length of the shorter? These are examples of probability problems in which the sample space is uncountable since it consists of all points on a line segment. This section extends the definition of probability to include uncountable sample spaces.

If we were to use the same procedure as for countable sample spaces we would start with an arbitrary uncountable set S and a Boolean o-algebra $\mathscr{B}$ of subsets of S and define a probability measure to be a completely additive nonnegative set function P defined on $\mathscr{B}$ with $P(S) = 1$. As it turns out, this procedure leads to certain technical difficulties that do not occur when S is countable. To attempt to describe these difficulties would take us too far afield. We shall avoid these difficulties by imposing restrictions at the outset on the set S and on the Boolean algebra $\mathscr{B}$.

First, we restrict S to be a subset of the real line **R,** or of n-space **R”.** For the Boolean algebra $\mathscr{B}$ we use special subsets of S which, in the language of modern integration theory, are called *measurable* subsets of S. We shall not attempt to describe the exact meaning of a measurable set; instead, we shall mention some of the properties possessed by the class of measurable sets.

First we consider subsets of **R.** The measurable subsets have the following properties:

1. If A is measurable, so is **R** $-$ A, the complement of A.
2. If {A,, *A,, A,,* . . .} is a countable collection of measurable sets, then the union *A,* $\cup$ *A,* $\cup$ *A,* $\cup \cdots$ is also measurable.
3. Every interval (open, closed, half-open, finite, or infinite) is measurable.

Thus, the measurable sets of **R** form a Boolean o-algebra which contains the intervals. A smallest Boolean o-algebra exists which has this property; its members are called ***Borel*** *sets,* after the French mathematician Émile Bore1 (1871-1956).

Similarly, in 2-space a smallest Boolean o-algebra exists which contains all Cartesian products of pairs of intervals; its members are the two-dimensional Bore1 sets. Bore1 sets in n-space are defined in an analogous fashion.

Henceforth, whenever we use a set S of real numbers as a sample space, or, more generally, whenever we use a set S in n-space as a sample space, we shall assume that this set is a Bore1 set. The Bore1 subsets of S themselves form a Boolean a-algebra. These subsets are extensive enough to include all the events that occur in the ordinary applications of probability theory.

DEFINITION OF PROBABILITY FOR UNCOUNTABLE SAMPLE SPACES. *Let S be a subset of $\mathbf{R}^n$, and let $\mathscr{B}$ be the Boolean a-algebra of Bore1 subsets of S. A nonnegative completely additive set function P defined on $\mathscr{B}$ with P(S) = 1 is called a probability measure. The triple (S, $\mathscr{B}$, P) is called a probability space.*

14.2 Countability of the set of points with positive probability

For countable sample spaces the probability of an event A is often computed by adding the point probabilities $P(x)$ for all x in A. This process is not fruitful for uncountable sample spaces because, as the next theorem shows, most of the point probabilities are zero.

THEOREM 14.1. *Let $(S, \mathscr{B}, P)$ be a probability space and let T denote the set of all x in S for which $P(x) > 0$. Then T is countable.*

Proof. For each $n = 1, 2, 3, \ldots$, let T_n denote the following subset of S:

$$T_n = \left\{x \,\middle|\, \frac{1}{n+1} < P(x) \le \frac{1}{n}\right\}.$$

If $P(x) > 0$ then $x \in T_n$ for some n. Conversely, if $x \in T_n$ for some n then $x \in T$. Hence $T = T_1 \cup T_2 \cup \cdots$. Now T_n contains at most n points, because if there were $n + 1$ or more points in T_n the sum of their point probabilities would exceed 1. Therefore T is countable, since it is a countable union of finite sets.

Theorem 14.1 tells us that positive probabilities can be assigned to at most a countable subset of S. The remaining points of S will have probability zero. In particular, if all the outcomes of S are equally likely, then every point in S must be assigned probability zero.

Note: Theorem 14.1 can be given a physical interpretation in terms of mass distribution which helps to illustrate its meaning. Imagine that we have an amount of mass, with the total quantity equal to 1. (This corresponds to P(S) = 1.) Suppose we are able to distribute this mass in any way we please along the real line, either by smearing it along the line with a uniform or perhaps a varying thickness, or by placing discrete lumps of mass at certain points, or both. (We interpret a positive amount of mass as a discrete lump.) We can place all the mass at one point. We can divide the mass equally or unequally in discrete lumps among two points, among ten points, among a million points, or among a countably infinite set of points. For example, we can put $\frac{1}{2}$ at 1, $\frac{1}{4}$ at 2, $\frac{1}{8}$ at 3, and so on, with mass $(\frac{1}{2})^n$ at each integer $n \ge 1$. Or we can smear all themasswithoutanyconcentrated lumps. Or we can smear part of it and distribute the rest in discrete lumps. Theorem 14.1 tells us that at most a countable set of points can be assigned discrete lumps of mass.

Since most (if not all) the point probabilities for an uncountable sample space will be zero, a knowledge of the point probabilities alone does not suffice to compute the probabilities of arbitrary events. Further information is required; it is best described in terms of two new concepts, *random variables* and *distribution functions,* to which we turn next. These concepts make possible the use of integral calculus in many problems with uncountable sample spaces. Integration takes the place of summation in the computation of probabilities.

14.3 Random variables

In many experiments we are interested in *numbers* associated with the outcomes of the experiment. For example, n coins are tossed simultaneously and we ask for the number of heads. A pair of dice is rolled and we ask for the sum of the points on the upturned faces. A dart is thrown at a circular target and we ask for its distance from the center. Whenever we associate a real number with each outcome of an experiment we are dealing with a *function* whose domain is the set of possible outcomes and whose range is the set of real numbers in question. Such a function is called a *random variable*. A formal definition can be given as follows:

DEFINITION OF A RANDOM VARIABLE. *Let S denote a sample space. A real-valuedfunction defined on S is called a one-dimensional random variable. If the function values are ordered pairs of real numbers (that is, vectors in 2-space), the function is said to be a two-dimensional random variable. More generally, an n-dimensional random variable is simply a function whose domain is the given sample space S and whose range is a collection of n-tuples of real numbers (vectors in n-space).*

Thus, a random variable is nothing but a vector-valued function defined on a set. The term "random" is used merely to remind us that the set in question is a sample space.7

Because of the generality of the above definition, it is possible to have many random variables associated with a given experiment. In any particular example the experimenter must decide which random variables will be of interest and importance to him. In general, we try to work with random variables whose function values reflect, as simply as possible, the properties of the outcomes of the experiment which are really essential.

Notations. Capital letters such as X, Y, Z are ordinarily used to denote one-dimensional random variables. A typical outcome of the experiment (that is, a typical element of the sample space) is usually denoted by the Greek letter ω (omega). Thus, $X(\omega)$ denotes that real number which the random variable X associates with the outcome ω.

The following are some simple examples of random variables.

EXAMPLE 1. An experiment consists of rolling a die and reading the number of points on the upturned face. The most "natural" random variable X to consider is the one stamped on the die by the manufacturer, namely:

$$X(w) = \omega \qquad \text{for } \omega = 1, 2, 3, 4, 5, 6.$$

† The terms "stochastic variable" and "chance variable" are also used as synonyms for "random variable." The word "stochastic" is derived from a Greek stem meaning "chance" and seems to have been invented by Jakob Bernoulli. It is commonly used in the literature of probability theory.

If we are interested in whether the number of points is even or odd, then we can consider instead the random variable Y, which is defined as follows:

$$Y(0) = 0 \quad \text{if } \omega \text{ is even,}$$

$$Y(w) = 1 \quad \text{if } w \text{ is odd.}$$

The values 0 and 1 are not essential-any two distinct real numbers could be used instead. However, 0 and 1 suggest "even" and "odd," respectively, because they represent the remainder obtained when the outcome ω is divided by 2.

EXAMPLE 2. A dart is thrown at a circular target. The set of all possible outcomes is the set of all points ω on the target. If we imagine a coordinate system placed on the target with the origin at the center, we can assign various random variables to this experiment. A natural one is the two-dimensional random variable which assigns to the point ω its rectangular coordinates (x, y). Another is that which assigns to ω its polar coordinates *(r,* 19). Examples of one-dimensional random variables are those which assign to each ω just one of its coordinates, such as the x-coordinate or the r-coordinate (distance from the origin). In an experiment of this type we often wish to know the probability that the dart will land in a particular region of the target, for example, the first quadrant. This event can be described most simply by the random variable which assigns to each point ω its polar coordinate angle θ, so that $X(\omega) = \theta$; the event "the dart lands in the first quadrant" is the set of ω such that $0 \leq X(\omega) \leq \frac{1}{2}\pi$.

Abbreviations. We avoid cumbersome notation by using special abbreviations to describe certain types of events and their probabilities. For example, if t is a real number, the set of all ω in the sample space such that $X(\omega) = t$ is denoted briefly by writing

$$X = t.$$

The probability of this event is written $P(X = t)$ instead of the lengthier $P(\{\omega \mid X(\omega) = t\})$. Symbols such as $P(X = a \text{ or } X = b)$ and $P(a < X \leq b)$ are defined in a similar fashion. Thus, the event "$X = a$ or $X = b$" is the union of the two events "$X = a$" and "$X = b$"; the symbol $P(X = a \text{ or } X = b)$ denotes the probability of this union. The event "$a < X \leq b$" is the set of all points ω such that $X(\omega)$ lies in the half-open interval $(a, b]$, and the symbol $P(a < X \leq b)$ denotes the probability of this event.

14.4 Exercises

1. Let X be a one-dimensional random variable.
   (a) If $a < b$, show that the two events $a < X \leq b$ and $X \leq a$ are disjoint.
   (b) Determine the union of the two events in part (a).
   (c) Show that $P(a < X \leq b) = P(X \leq b) - P(X \leq a)$.
2. Let (X, Y) denote a two-dimensional random variable defined on a sample space S. This means that (X, Y) is a function which assigns to each ω in S a pair of real numbers $(X(\omega), Y(\omega))$. Of course, each of X and Y is a one-dimensional random variable defined on S. The notation

$$X \leq a, \; Y \leq b$$

stands for the set of all elements ω in S such that $X(\omega) \leq a$ and $Y(\omega) \leq b$.

(a) If $a < b$ and $c < d$, describe, in terms of elements of S, the meaning of the following notation: $a < X \leq b, c < Y \leq d$.

(b) Show that the two events "$X \leq a, Y \leq c$" and "$X \leq a, c < Y \leq d$" are disjoint. Interpret these events geometrically.

(c) Determine the union of the two events in (b).

(d) Generalize Exercise 1 (c) to the two-dimensional case.

3. Two fair dice are rolled, each outcome being an ordered pair (a, b), where each of a and b is an integer from 1 to 6. Let X be the random variable which assigns the value $a + b$ to the outcome (a, b).
(a) Describe, in roster notation, the events "$X = 7$," "$X = 11$," "$X = 7$ or $X = 11$."
(b) Compute the probabilities of the events in part (a).
4. Consider an experiment in which four coins are tossed simultaneously (or one coin is tossed four times). For each coin define a random variable which assigns the value 1 to heads and the value 0 to tails, and denote these random variables by X_1, X_2, X_3, X_4. Assign the probabilities $P(X_i = 1) = P(X_i = 0) = \frac{1}{2}$ for each X_i. Consider a new random variable Y which assigns to each outcome the total number of heads among the four coins. Express Y in terms of X_1, X_2, X_3, X_4 and compute the probabilities $P(Y = 0)$, $P(Y = 1)$, and $P(Y \leq 1)$.
5. A small railroad company has facilities for transporting 100 passengers a day between two cities, at a fixed cost (to the company) of \$7 per passenger. If more than 100 passengers buy tickets in any one day the railroad is obligated to provide bus transportation for the excess at a cost of \$10 per passenger. Let X be the random variable which counts the number of passengers that buy tickets in a given day. The possible values of X are the integers 0, 1, 2, 3, . . . up to a certain unknown maximum. Let Y denote the random variable which describes the total daily cost (in dollars) to the railroad for handling passengers. Express Y in terms of X.
6. A factory production line consists of two work stations A and B. At station A, X units per hour are assembled; they are immediately transported to station B, where they are inspected at the rate of Y units per hour, where $Y < X$. The possible values of X and Y are the integers 8, 9, and 10. Let Z denote the random variable which counts the number of units that come off the production line during the first hour of production.
(a) Express Z in terms of X and Y, assuming each of X and Y is constant during this hour.
(b) Describe, in a similar way, the random variable U which counts the number of units delivered in the first two consecutive hours of production. Each of X and Y is constant during each hour, but the constant values during the second hour need not be the same as those during the first.

14.5 Distribution functions

We turn now to the problem of computing the probabilities of events associated with a given random variable. Let X be a one-dimensional random variable defined on a sample space S, where S is a Borel set in n-space for some $n \geq 1$. Let P be a probability measure defined on the Borel subsets of S. For each ω in S, $X(\omega)$ is a real number, and as ω runs through the elements of S the numbers $X(\omega)$ run through a set of real numbers (the range of X). This set may be finite, countably infinite, or uncountable. For each real number t we consider the following special subset of S:

$$A(t) = \{\omega \mid X(\omega) \leq t\}.$$

If t is less than all the numbers in the range of X, the set $A(t)$ will be empty; otherwise, $A(t)$ will be a nonempty subset of S. We assume that for each t the set $A(t)$ is an ***event***, that is, a Borel set. According to the convention discussed at the end of Section 14.3, we denote this event by the symbol $X \le t$.

Suppose we know the probability $P(X \le t)$ for every real t. We shall find in a moment that this knowledge enables us to compute the probabilities of many other events of interest. This is done by using the probabilities $P(X \le t)$ as a basis for constructing a new function F, called the *distribution function* of X. It is defined as follows:

DEFINITION OF A DISTRIBUTION FUNCTION. *Let X be a one-dimensional random variable. The function F defined for all real t by the equation*

$$F(t) = P(X \le t)$$

is called the distribution function of the random variable X.

> *Note:* Sometimes the notation F_X is used to emphasize the fact that the distribution function is associated with the particular random variable X. The value of the function at t is then denoted by $F_X(t)$.

It is important to realize that the distribution function F is defined over the entire real axis, even though the range of X may be only a bounded portion of the real axis. In fact, if all numbers $X(\omega)$ lie in some finite interval $[a, b]$, then for $t < a$ the probability $P(X \le t)$ is zero (since for $t < a$ the set $X \le t$ is empty) and for $t \ge b$ the probability $P(X \le t)$ is 1 (because in this case the set $X \le t$ is the entire sample space). This means that for *bounded* random variables X whose range is within an interval $[a, b]$ we have $F(t) = 0$ for all $t < a$ and $F(t) = 1$ for all $t \ge b$.

We now proceed to derive a number of properties common to all distribution functions.

THEOREM 14.2. *Let F denote a distribution function of a one-dimensional random variable X. Then we have:*

(a) $0 \le F(t) \le 1$ *for all* t.

(b) $P(a < X \le b) = F(b) - F(a)$ *if* $a < b$.

(c) $F(a) \le F(b)$ *if* $a < b$.

Proof. Part (a) follows at once from the definition of F because probabilities always lie between 0 and 1.

To prove (b) we note that the events "$a < X \le b$" and "$X \le a$" are disjoint. Their union is the event "$X \le b$." Using additivity we obtain

$$P(a < X \le b) + P(X \le a) = P(X \le b),$$

which can also be expressed as

$$P(a < X \le b) = P(X \le b) - P(X \le a) = F(b) - F(a).$$

Part (c) follows from (b) since $P(a < X \le b) \ge 0$.

> *Note:* Using the mass analogy, we would say that $F(t)$ represents the total amount of mass located between $-\infty$ and t (including the point t itself). The amount of mass located in a half-open interval $(a, b]$ is $F(b) - F(a)$.

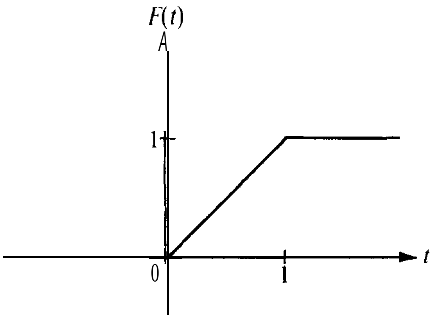


FIGURE 14.1 A distribution function of a bounded random variable.

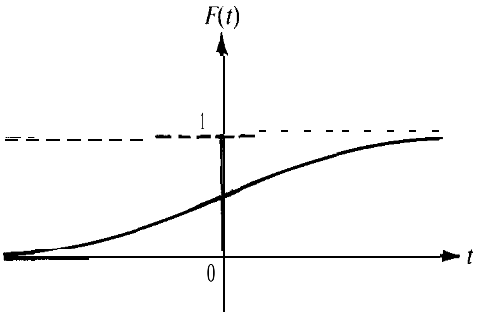


FIGURE 14.2 A distribution function of an unbounded random variable.

Figure 14.1 shows a distribution function of a bounded random variable X whose values $X(\omega)$ lie in the interval [0, 1]. This particular example is known as a *uniform distribution.* Here we have

$$F(t) = 0 \text{ for } t < 0, \quad F(t) = t \text{ for } 0 \le t \le 1, \quad F(t) = 1 \text{ for } t \ge 1.$$

Figure 14.2 shows an example of a distribution function corresponding to an unbounded random variable. This example is known as a *Cauchy distribution* and its function values are given by the formula

$$F(t) = \frac{1}{2} + \frac{1}{\pi} \arctan t .$$

Experiments that lead to uniform and to Cauchy distributions will be discussed later.

Note: Using the mass analogy, we would say that in Figure 14.1 no mass has been placed to the left of the origin or to the right of the point 1. The entire mass has been distributed over the interval [0, 1]. The graph of F is linear over this interval because the mass is smeared with a uniform thickness. In Figure 14.2 the mass has been smeared along the entire axis. The graph is nonlinear because the mass has been smeared with a varying thickness.

Theorem 14.2(b) tells us how to compute (in terms of F) the probability that X lies in a half-open interval of the form $(a, b]$. The next theorem deals with other types of intervals.

THEOREM 14.3. *Let F be a distribution function of a one-dimensional random variable X. Then if $a < b$ we have:*

(a) $P(a \le X \le b) = F(b) - F(a) + P(X = a)$.
(b) $P(a < X < b) = F(b) - F(a) - P(X = b)$.
(c) $P(a \le X < b) = F(b) - F(a) + P(X = a) - P(X = b)$.

Proof. To prove (a) we note that the events "$a < X \le b$" and "$X = a$" are disjoint and their union is "$a \le X \le b$." Using additivity and Theorem 14.2(b) we obtain (a). Parts (b) and (c) are similarly proved.

Note that all four events

$$a < X \le b, \qquad a \le X \le b, \qquad a < X < b, \qquad \text{and} \qquad a \le X < b$$

have equal probabilities if and only if $P(X = a) = 0$ and $P(X = b) = 0$.

The examples shown in Figures 14.1 and 14.2 illustrate two further properties shared by all distribution functions. They are described in the following theorem.

THEOREM **14.4.** *Let F be the distribution function of a one-dimensional random variable X. Then we have*

(14.1) $$\lim_{t \to -\infty} F(t) = 0 \quad \textit{and} \quad \lim_{t \to +\infty} F(t) = 1 .$$

Proof. The existence of the two limits in (14.1) and the fact that each of the two limits lies between 0 and 1 follow at once, since F is a monotonic function whose values lie between 0 and 1.

Let us denote the limits in (14.1) by L_1 and L_2, respectively. To prove that $L_1 = 0$ and that $L_2 = 1$ we shall use the countably additive property of probability. For this purpose we express the whole space S as a countable union of disjoint events:

$$S = \bigcup_{n=1}^{\infty} (-n < X \le -n + 1) \cup \bigcup_{n=0}^{\infty} (n < X \le n + 1).$$

Then, using additivity, we get

$$\begin{aligned} P(S) &= \sum_{n=1}^{\infty} P(-n < X \le -n + 1) + \sum_{n=0}^{\infty} P(n < X \le n + 1) \\ &= \lim_{M \to \infty} \sum_{n=1}^{M} [F(-n + 1) - F(-n)] + \lim_{N \to \infty} \sum_{n=0}^{N} [F(n + 1) - F(n)]. \end{aligned}$$

The sums on the right will telescope, giving us

$$\begin{aligned} P(S) &= \lim_{M \to \infty} [F(0) - F(-M)] + \lim_{N \to \infty} [F(N + 1) - F(0)] \\ &= F(0) - L_1 + L_2 - F(0) = L_2 - L_1 . \end{aligned}$$

Since $P(S) = 1$, this proves that $L_2 - L_1 = 1$ or $L_2 = 1 + L_1$. On the other hand, we also have $L_2 \le 1$ and $L_1 \ge 0$. This implies that $L_1 = 0$ and $L_2 = 1$, as asserted.

14.6 Discontinuities of distribution functions

An example of a possible distribution function with discontinuities is shown in Figure 14.3. Using the mass analogy we would say that F has a jump discontinuity at each point which carries a positive amount of mass. As the next theorem shows, the jump is equal to the amount of mass concentrated at that particular point.

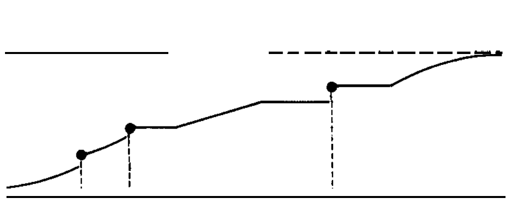


FIGURE 14.3 A possible distribution function.

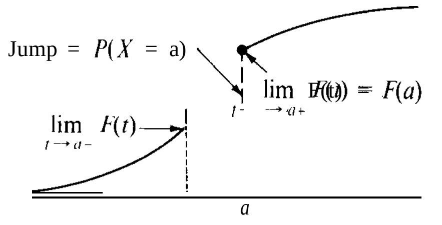


FIGURE 14.4 Illustrating a jump discontinuity of a distribution function.

THEOREM 14.5. *Let F be the distribution function of a one-dimensional random variable X. Then for each real a we have*

$$\lim_{t \to a+} F(t) = F(a) \tag{14.2}$$

and

$$\lim_{t \to a-} F(t) = F(a) - P(X = a). \tag{14.3}$$

Note: The limit relation in (14.2) tells us that F is *continuous from the right* at each point a, because $F(t) \to F(a)$ as $t \to a$ from the right. On the other hand, Equation (14.3) tells us that as $t \to a$ from the left, $F(t)$ will approach $F(a)$ if and only if the probability $P(X = a)$ is zero. When $P(X = a)$ is not zero, the graph of F has a jump discontinuity at a of the type shown in Figure 14.4.

Proof. The existence of the limits follows at once from the monotonicity and boundedness of F. We prove now that the limits have the values indicated. For this purpose we use part (b) of Theorem 14.2. If $t > a$ we write

$$F(t) = F(a) + P(a < X \le t)\,; \tag{14.4}$$

if $t < a$ we write

$$F(t) = F(a) - P(t < X \le a). \tag{14.5}$$

Letting $t \to a+$ in (14.4) we find

$$\lim_{t \to a+} F(t) = F(a) + \lim_{t \to a+} P(a < X \le t)\,,$$

whereas if $t \to a-$ in (14.5) we obtain

$$\lim_{t \to a-} F(t) = F(a) - \lim_{t \to a-} P(t < X \le a).$$

Therefore to prove (14.2) and (14.3) we must establish two equations:

$$\lim_{t \to a+} P(a < X \le t) = 0 \tag{14.6}$$

and

(14.7)
$$\lim_{t\to a-} P(t < X \le a) = P(X = a).$$

These can be justified intuitively as follows: When $t \to a+$, the half-open interval $(a, t]$ shrinks to the empty set. That is, the intersection of all half-open intervals $(a, t]$, for $t > a$, is empty. On the other hand, when $t \to a-$ the half-open interval $(t, a]$ shrinks to the point a. (The intersection of all intervals $(t, a]$ for $t < a$ is the set $\{a\}$.) Therefore, if probability behaves in a continuous fashion, Equations (14.6) and (14.7) must be valid. To convert this argument into a rigorous proof we proceed as follows:

For each integer $n \ge 1$, let

(14.8)
$$p_n = P\left(a < X \le a + \frac{1}{n}\right).$$

To prove (14.6) it suffices to show that $p_n \to 0$ as $n \to \infty$. Let S_n denote the event

$$a + \frac{1}{n+1} < X \le a + \frac{1}{n}.$$

The sets S_n are disjoint and their union $S_1 \cup S_2 \cup S_3 \cup \cdots$ is the event $a < X \le a + 1$. By countable additivity we have

(14.9)
$$\sum_{n=1}^{\infty} P(S_n) = P(a < X \le a + 1) = p_1.$$

On the other hand, Equation (14.8) implies that

$$p_n - p_{n+1} = P(S_n),$$

so from (14.9) we obtain the relation

(14.10)
$$\sum_{n=1}^{\infty} (p_n - p_{n+1}) = p_1.$$

The convergence of the series is a consequence of (14.9). But the series on the left of (14.10) is a telescoping series with sum

$$p_1 - \lim_{n\to\infty} p_n.$$

Therefore (14.10) implies that $\lim_{n\to\infty} p_n = 0$, and this proves (14.6).

A slight modification of this argument enables us to prove (14.7) as well. Since

$$P(t < X \le a) = P(t < X < a) + P(X = a)$$

we need only prove that

$$\lim_{t\to a-} P(t < X < a) = 0.$$

For this purpose we introduce the numbers

$$q_n = P\left(a - \frac{1}{n} < X < a\right)$$

and show that $q_n \to 0$ as $n \to \infty$. In this case we consider the events T_n given by

$$a - \frac{1}{n} < X \le a - \frac{1}{n+1}$$

for $n = 1, 2, 3, \ldots$. These are disjoint and their union is the event $a - 1 < X < a$, so we have

$$\sum_{n=1}^{\infty} P(T_n) = P(a - 1 < X < a) = q_1 .$$

We now note that $q_n - q_{n+1} = P(T_n)$, and we complete the proof as above.

The most general type of distribution is any real-valued function F that has the following properties :

(a) F is monotonically increasing on the real axis,
(b) F is continuous from the right at each point,
(c) $\lim_{t\to-\infty} F(t) = 0$ and $\lim_{t\to+\infty} F(t) = 1$.

In fact, it can be shown that for each such function F there is a corresponding set function P, defined on the Bore1 sets of the real line, such that P is a probability measure which assigns the probability $F(b) - F(a)$ to each half-open interval $(a, b]$. For a proof of this statement, see H. Cramer, *Mathematical Methods of Statistics*, Princeton University Press, Princeton, N.J., 1946.

There are two special types of distributions, known as *discrete* and *continuous*, that are of particular importance in practice. In the discrete case the entire mass is concentrated at a finite or countably infinite number of points, whereas in the continuous case the mass is smeared, in uniform or varying thickness, along an interval (finite or infinite). These two types of distributions will be treated in detail in the next few sections.

14.7 Discrete distributions. Probability mass functions

Let X be a one-dimensional random variable and consider a new function p, called the *probability mass function* of X. Its values $p(t)$ are defined for every real number t by the equation

$$p(t) = P(X = t).$$

That is, $p(t)$ is the probability that X takes the value t. When we want to emphasize that p is associated with X we writep, instead ofp and $p_X(t)$ instead of $p(t)$.

The set of real numbers t for which $p(t) > 0$ is either finite or countable. We denote this set by T; that is, we let

$$T = \{t \mid p(t) > 0\}.$$

The random variable X is said to be ***discrete*** if

$$\sum_{t\in T} p(t) = 1 .$$

In other words, X is discrete if a unit probability mass is distributed over the real line by concentrating a positive mass $p(t)$ at each point t of some finite or countably infinite set T and no mass at the remaining points. The points of T are called the ***mass points*** of X.

For discrete random variables a knowledge of the probability mass function enables us to compute probabilities of arbitrary events. In fact, we have the following theorem.

THEOREM 14.6. ***Let A be a Borel subset of the real line*** R, ***and let $P(X \in A)$ denote the probability of the set of ω such that $X(\omega) \in A$. Then we have***

$$P(X \in A) = \sum_{x\in A\cap T} p(x) , \tag{14.11}$$

where T is the set of mass points of X.

Proof. Since $A \cap T \subseteq A$ and $T - A \subseteq R - A$, we have

$$\sum_{x\in A\cap T} p(x) \le P(X \in A) \quad \text{and} \quad \sum_{x\in T-A} p(x) \le P(X \notin A). \tag{14.12}$$

But $A \cap T$ and $T - A$ are disjoint sets whose union is T, so the second inequality in (14.12) is equivalent to

$$1 - \sum_{x\in A\cap T} p(x) \le 1 - P(X \in A) \qquad \text{or} \qquad \sum_{x\in A\cap T} p(x) \ge P(X \in A).$$

Combining this with the first inequality in (14.12) we obtain (14.11).

Note: Since $p(x) = 0$ when $x \notin T$, the sum on the right of (14.11) can be written as $\sum_{x\in A} p(x)$ without danger of its being misunderstood.

When A is the interval $(-\infty, t]$, the sum in (14.11) gives the value of the distribution function $F(t)$. Thus, we have

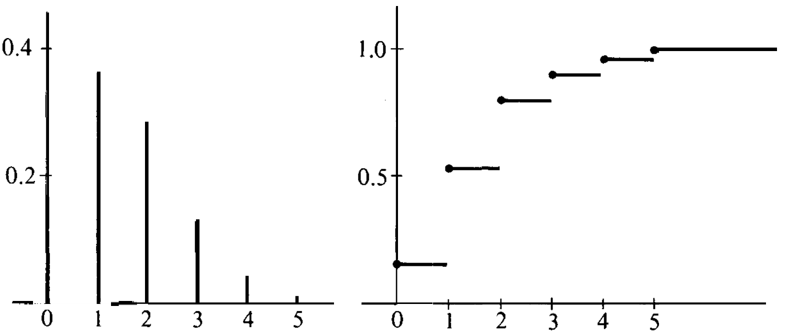
$$F(t) = P(X \le t) = \sum_{x\le t} p(x) .$$

If a random variable X is discrete, the corresponding distribution function F is also called discrete.

The following examples of discrete distributions occur frequently in practice.

EXAMPLE 1. ***Binomial distribution.*** Let p be a given real number satisfying $0 \le p \le 1$ and let $q = 1 - p$. Suppose a random variable X assumes the values 0, 1, 2, . . . , n, where n is a fixed positive integer, and suppose the probability $P(X = k)$ is given by the formula

$$P(X = k) = \binom{n}{k} p^k q^{n-k} \qquad \text{for} \quad k = 0, 1, 2, \ldots, n.$$



(a) The mass function (b) The distribution function

FIGURE 14.5 The probability mass function and the distribution function of a binomial distribution with parameters $n = 5$ and $p = \frac{1}{3}$.

This assignment of probabilities is permissible because the sum of all the point probabilities is

$$\sum_{k=0}^{n} P(X = k) = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} = (p + q)^n = 1.$$

The corresponding distribution function F_X is said to be a *binomial distribution* with parameters n andp. Its values may be computed by the summation formula

$$F_X(t) = \sum_{0 \le k \le t} \binom{n}{k} p^k q^{n-k}.$$

Binomial distributions arise naturally from a Bernoulli sequence of trials where p is the probability of "success" and q the probability of "failure." In fact, when the random variable X counts the number of successes in n trials, $P(X = k)$ is precisely $\binom{n}{k} p^k q^{n-k}$ because of Bernoulli's formula. (See Theorem 13.3 in Section 13.16.) Figure 14.5 shows the graphs of the probability mass function and the corresponding distribution function for a binomial distribution with parameters $n = 5$ and $p = \frac{1}{3}$.

EXAMPLE 2. *Poisson distribution*. Let λ be a positive real number and let a random variable X assume the values 0, 1, 2, 3, ... If the probability $P(X = k)$ is given by the formula

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!} \qquad \text{for} \quad k = 0, 1, 2, \ldots,$$

the corresponding distribution function F_X is said to be a *Poisson distribution* with parameter λ. It is so named in honor of the French mathematician S. D. Poisson (1781-1840). This assignment of probabilities is permissible because

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

The values of the distribution function are computed from the partial sums

$$F_X(t) = e^{-\lambda} \sum_{0 \le k \le t} \frac{\lambda^k}{k!}.$$

The Poisson distribution is applicable to many problems involving random events occurring in time, such as traffic accidents, connections to wrong numbers in a telephone exchange, and chromosome interchanges in cells induced by x-ray radiation. Some specific applications are discussed in the books by Feller and Parzen listed at the end of this chapter.

14.8 Exercises

1. A perfectly balanced die is rolled. For a random variable X we take the function which counts the number of points on the upturned face. Draw a graph of the corresponding distribution function F_X.
2. Two dice are rolled. Let X denote the random variable which counts the total number of points on the upturned faces. Construct a table giving the nonzero values of the probability mass function p_X and draw a graph of the corresponding distribution function F_X.
3. The distribution function F of a random variable X is given by the following formulas:

$$F(t) = \begin{cases} 0 & \text{if } t < -2, \\ \frac{1}{2} & \text{if } -2 \le t < 0, \\ 4 & \text{if } 0 \le t < 2, \\ 1 & \text{if } t \ge 2. \end{cases}$$

   (a) Sketch the graph of F.
   (b) Describe the probability mass function p and draw its graph.
   (c) Compute the following probabilities: $P(X = 1)$, $P(X \le 1)$, $P(X < 1)$, $P(X = 2)$, $P(X \le 2)$, $P(0 < X < 2)$, $P(0 < X \le 2)$, $P(1 \le X \le 2)$.
4. Consider a random variable X whose possible values are all rational numbers of the form $\frac{n}{n+1}$ and $\frac{n+1}{n}$, where $n = 1, 2, 3, \ldots$. If

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

   verify that this assignment of probabilities is permissible and sketch the general shape of the graph of the distribution function F”.
5. The probability mass function p of a random variable X is zero except at points $t = 0, 1, 2$. At these points it has the values

$$p(0) = 3c^3, \qquad p(1) = 4c - 10c^2, \qquad p(2) = 5c - 1,$$

   for some $c > 0$.
   (a) Determine the value of c.
   (b) Compute the following probabilities: $P(X < 1)$, $P(X < 2)$, $P(1 < X \le 2)$, $P(0 < X < 3)$.
   (c) Describe the distribution function F and sketch its graph.
   (d) Find the largest t such that $F(t) < \frac{1}{2}$.
   (e) Find the smallest t such that $F(t) > \frac{1}{3}$.