1

Properties of Probability Distributions

1.1 Introduction

Distribution theory is concerned with probability distributions of random variables, with the emphasis on the types of random variables frequently used in the theory and application of statistical methods. For instance, in a statistical estimation problem we may need to determine the probability distribution of a proposed estimator or to calculate probabilities in order to construct a confidence interval.

Clearly, there is a close relationship between distribution theory and probability theory; in some sense, distribution theory consists of those aspects of probability theory that are often used in the development of statistical theory and methodology. In particular, the problem of deriving properties of probability distributions of statistics, such as the sample mean or sample standard deviation, based on assumptions on the distributions of the underlying random variables, receives much emphasis in distribution theory.

In this chapter, we consider the basic properties of probability distributions. Although these concepts most likely are familiar to anyone who has studied elementary probability theory, they play such a central role in the subsequent chapters that they are presented here for completeness.

1.2 Basic Framework

The starting point for probability theory and, hence, distribution theory is the concept of an *experiment*. The term experiment may actually refer to a physical experiment in the usual sense, but more generally we will refer to something as an experiment when it has the following properties: there is a well-defined set of possible outcomes of the experiment, each time the experiment is performed exactly one of the possible outcomes occurs, and the outcome that occurs is governed by some chance mechanism.

Let Ω denote the *sample space* of the experiment, the set of possible outcomes of the experiment; a subset A of Ω is called an *event*. Associated with each event A is a probability $\mathrm{P}(A)$. Hence, P is a function defined on subsets of Ω and taking values in the interval $[0, 1]$. The function P is required to have certain properties:

(P1) $\mathrm{P}(\Omega) = 1$
(P2) If A and B are disjoint subsets of Ω, then $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B)$.

(P3) If $A_1, A_2, \ldots,$ are disjoint subsets of Ω, then

$$\mathrm{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathrm{P}(A_n).$$

Note that (P3) implies (P2); however, (P3), which is concerned with an infinite sequence of events, is of a different nature than (P2) and it is useful to consider them separately. There are a number of straightforward consequences of (P1)–(P3). For instance, $\mathrm{P}(\emptyset) = 0$, if A^c denotes the complement of A, then $\mathrm{P}(A^c) = 1 - \mathrm{P}(A)$, and, for A_1, A_2 not necessarily disjoint,

$$\mathrm{P}(A_1 \cup A_2) = \mathrm{P}(A_1) + \mathrm{P}(A_2) - \mathrm{P}(A_1 \cap A_2).$$

Example* 1.1** ***(Sampling from a finite population). Suppose that Ω is a finite set and that, for each $\omega \in \Omega$,

$$\mathrm{P}(\{\omega\}) = c$$

for some constant c. Clearly, $c = 1/|\Omega|$ where $|\Omega|$ denotes the cardinality of Ω.

Let A denote a subset of Ω. Then

$$\mathrm{P}(A) = \frac{|A|}{|\Omega|}.$$

Thus, the problem of determining $\mathrm{P}(A)$ is essentially the problem of counting the number of elements in A and Ω. □

Example* 1.2** ***(Bernoulli trials). Let

$$\Omega = \{x \in \mathbf{R}^n \colon x = (x_1, \ldots, x_n), x_j = 0 \text{ or } 1, \quad j = 1, \ldots, n\}$$

so that an element of Ω is a vector of ones and zeros. For $\omega = (x_1, \ldots, x_n) \in \Omega$, take

$$\mathrm{P}(\omega) = \prod_{j=1}^{n} \theta^{x_j} (1-\theta)^{1-x_j}$$

where $0 < \theta < 1$ is a given constant. □

Example* 1.3** ***(Uniform distribution). Suppose that $\Omega = (0, 1)$ and suppose that the probability of any interval in Ω is the length of the interval. More generally, we may take the probability of a subset A of Ω to be

$$\mathrm{P}(A) = \int_A dx.$$ □

Ideally, P is defined on the set of all subsets of Ω. Unfortunately, it is not generally possible to do so and still have properties (P1)–(P3) be satisfied. Instead P is defined only on a set $\mathcal{F}$ of subsets of Ω; if $A \subset \Omega$ is not in $\mathcal{F}$, then $\mathrm{P}(A)$ is not defined. The sets in $\mathcal{F}$ are said to be *measurable*. The triple $(\Omega, \mathcal{F}, \mathrm{P})$ is called a *probability space*; for example, we might refer to a random variable X defined on some probability space.

Clearly for such an approach to probability theory to be useful for applications, the set $\mathcal{F}$ must contain all subsets of Ω of practical interest. For instance, when Ω is a countable set, $\mathcal{F}$ may be taken to be the set of all subsets of Ω. When Ω may be taken to be a

Euclidean space $\mathbf{R}^d$, $\mathcal{F}$ may be taken to be the set of all subsets of $\mathbf{R}^d$ formed by starting with a countable set of rectangles in $\mathbf{R}^d$ and then performing a countable number of set operations such as intersections and unions. The same approach works when Ω is a subset of a Euclidean space.

The study of theses issues forms the branch of mathematics known as measure theory. In this book, we avoid such issues and implicitly assume that any event of interest is measurable.

Note that condition (P3), which deals with an infinite number of events, is of a different nature than conditions (P1) and (P2). This condition is often referred to as *countable additivity* of a probability function. However, it is best understood as a type of continuity condition on P. It is easier to see the connection between (P3) and continuity if it is expressed in terms of one of two equivalent conditions. Consider the following:

(P4) If $A_1, A_2, \ldots$, are subsets of Ω satisfying $A_1 \subset A_2 \subset \cdots$, then

$$\mathrm{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mathrm{P}(A_n)$$

(P5) If $A_1, A_2, \ldots$, are subsets of Ω satisfying $A_1 \supset A_2 \supset \cdots$, then

$$\mathrm{P}\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mathrm{P}(A_n).$$

Suppose that, as in (P4), $A_1, A_2, \ldots$ is a sequence of increasing subsets of Ω. Then we may take the limit of this sequence to be the union of the A_n; that is,

$$\lim_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} A_n.$$

Condition (P4) may then be written as

$$\mathrm{P}\left(\lim_{n\to\infty} A_n\right) = \lim_{n\to\infty} \mathrm{P}(A_n).$$

A similar interpretation applies to (P5). Thus, (P4) and (P5) may be viewed as continuity conditions on P.

The equivalence of (P3), (P4), and (P5) is established in the following theorem.

***Theorem* 1.1.** *Consider an experiment with sample space Ω. Let* P *denote a function defined on subsets of Ω such that conditions* (P1) *and* (P2) *are satisfied. Then conditions* (P3), (P4), *and* (P5) *are equivalent in the sense that if any one of these conditions holds, the other two hold as well.*

Proof. First note that if $A_1, A_2, \ldots$ is an increasing sequence of subsets of Ω, then $A_1^c, A_2^c, \ldots$ is a decreasing sequence of subsets and, since, for each $k = 1, 2, \ldots$,

$$\left(\bigcup_{n=1}^{k} A_n\right)^c = \bigcap_{n=1}^{k} A_n^c,$$

$$\left(\lim_{n\to\infty} A_n\right)^c = \bigcap_{n=1}^{\infty} A_n^c = \lim_{n\to\infty} A_n^c.$$

Suppose (P5) holds. Then

$$\mathrm{P}\left(\lim_{n\to\infty} A_n^c\right) = \lim_{n\to\infty} \mathrm{P}\left(A_n^c\right)$$

so that

$$\mathrm{P}\left(\lim_{n\to\infty} A_n\right) = 1 - \mathrm{P}\left\{\left(\lim_{n\to\infty} A_n\right)^c\right\} = 1 - \lim_{n\to\infty} \mathrm{P}\left(A_n^c\right) = \lim_{n\to\infty} \mathrm{P}(A_n),$$

proving (P4). A similar argument may be used to show that (P4) implies (P5). Hence, it suffices to show that (P3) and (P4) are equivalent.

Suppose $A_1, A_2, \ldots$ is an increasing sequence of events. For $n = 2, 3, \ldots$, define

$$\bar{A}_n = A_n \cap A_{n-1}^c.$$

Then, for $1 < n < k$,

$$\bar{A}_n \cap \bar{A}_k = (A_n \cap A_k) \cap \left(A_{n-1}^c \cap A_{k-1}^c\right).$$

Note that, since the sequence $A_1, A_2, \ldots$ is increasing, and $n < k$,

$$A_n \cap A_k = A_n$$

and

$$A_{n-1}^c \cap A_{k-1}^c = A_{k-1}^c.$$

Hence, since $A_n \subset A_{k-1}$,

$$\bar{A}_n \cap \bar{A}_k = A_n \cap A_{k-1}^c = \emptyset.$$

Suppose $\omega \in A_k$. Then either $\omega \in A_{k-1}$ or $\omega \in A_{k-1}^c \cap A_k = \bar{A}_k$; similarly, if $\omega \in A_{k-1}$ then either $\omega \in A_{k-2}$ or $\omega \in A_1^c \cap A_{k-1} \cap A_{k-2}^c = \bar{A}_{k-1}$. Hence, ω must be an element of either one of $\bar{A}_k, \bar{A}_{k-1}, \ldots, \bar{A}_2$ or of A_1. That is,

$$A_k = A_1 \cup \bar{A}_2 \cup \bar{A}_3 \cup \cdots \cup \bar{A}_k;$$

hence, taking $\bar{A}_1 = A_1$,

$$A_k = \bigcup_{n=1}^{k} \bar{A}_n$$

and

$$\lim_{k\to\infty} A_k = \bigcup_{n=1}^{\infty} \bar{A}_n.$$

Now suppose that (P3) holds. Then

$$\mathrm{P}(\lim_{k\to\infty} A_k) = \mathrm{P}\left(\bigcup_{n=1}^{\infty} \bar{A}_n\right) = \sum_{n=1}^{\infty} \mathrm{P}(\bar{A}_n) = \lim_{k\to\infty} \sum_{n=1}^{k} \mathrm{P}(\bar{A}_n) = \lim_{k\to\infty} \mathrm{P}(A_k),$$

proving (P4).

Now suppose that (P4) holds. Let $A_1, A_2, \ldots$ denote an arbitrary sequence of disjoint subsets of Ω and let

$$A_0 = \bigcup_{n=1}^{\infty} A_n.$$

Define

$$\tilde{A}_k = \bigcup_{n=1}^{k} A_j, \quad k = 1, 2, \ldots;$$

note that $\tilde{A}_1, \tilde{A}_2, \ldots$ is an increasing sequence and that

$$A_0 = \lim_{k \to \infty} \tilde{A}_k.$$

Hence, by (P4),

$$\mathrm{P}(A_0) = \lim_{k \to \infty} \mathrm{P}(\tilde{A}_k) = \lim_{k \to \infty} \sum_{n=1}^{k} \mathrm{P}(A_n) = \sum_{n=1}^{\infty} \mathrm{P}(A_n),$$

proving (P3). It follows that (P3) and (P4) are equivalent, proving the theorem. ■

1.3 Random Variables

Let ω denote the outcome of an experiment; that is, let ω denote an element of Ω. In many applications we are concerned primarily with certain numerical characteristics of ω, rather than with ω itself. Let $X : \Omega \to \mathcal{X}$, where $\mathcal{X}$ is a subset of $\mathbf{R}^d$ for some $d = 1, 2, \ldots$, denote a *random variable*; the set $\mathcal{X}$ is called the *range* of X or, sometimes, the *sample space of* X. For a given outcome $\omega \in \Omega$, the corresponding value of X is $x = X(\omega)$. Probabilities regarding X may be obtained from the probability function P for the original experiment. Let P_X denote a function such that for any set $A \subset \mathcal{X}$, $\mathrm{P}_X(A)$ denotes the probability that $X \in A$. Then P_X is a probability function defined on subsets of $\mathcal{X}$ and

$$\mathrm{P}_X(A) = \mathrm{P}(\{\omega \in \Omega\colon X(\omega) \in A\}).$$

We will generally use a less formal notation in which $\Pr(X \in A)$ denotes $\mathrm{P}_X(A)$. For instance, the probability that $X \leq 1$ may be written as either $\Pr(X \leq 1)$ or $\mathrm{P}_X\{(-\infty, 1]\}$. In this book, we will generally focus on probabilities associated with random variables, without explicit reference to the underlying experiments and associated probability functions.

Note that since P_X defines a probability function on the subsets of $\mathcal{X}$, it must satisfy conditions (P1)–(P3). Also, the issues regarding measurability discussed in the previous section apply here as well.

When the range $\mathcal{X}$ of a random variable X is a subset of $\mathbf{R}^d$ for some $d = 1, 2, \ldots$, it is often convenient to proceed as if probability function P_X is defined on the entire space $\mathbf{R}^d$. Then the probability of any subset of $\mathcal{X}^c$ is 0 and, for any set $A \subset \mathbf{R}^d$,

$$\mathrm{P}_X(A) \equiv \Pr(X \in A) = \Pr(X \in A \cap \mathcal{X}).$$

It is worth noting that some authors distinguish between random variables and random vectors, the latter term referring to random variables X for which $\mathcal{X}$ is a subset of $\mathbf{R}^d$ for $d > 1$. Here we will not make this distinction. The term *random variable* will refer to either a scalar or vector; in those cases in which it is important to distinguish between real-valued and vector random variables, the terms *real-valued random variable* and *scalar random variable* will be used to denote a random variable with $\mathcal{X} \subset \mathbf{R}$ and the term *vector random variable* and *random vector* will be used to denote a random variable with $X \subset \mathbf{R}^d$, $d > 1$. Random vectors will always be taken to be column vectors so that a d-dimensional random

vector X is of the form

$$X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_d \end{pmatrix}$$

where $X_1, X_2, \ldots, X_d$ are real-valued random variables.

For convenience, when writing a d-dimensional random vector in the text, we will write $X = (X_1, \ldots, X_d)$ rather than $X = (X_1, \ldots, X_d)^T$. Also, if X and Y are both random vectors, the random vector formed by combining X and Y will be written as (X, Y), rather than the more correct, but more cumbersome, $(X^T, Y^T)^T$. We will often consider random vectors of the form (X, Y) with range $\mathcal{X} \times \mathcal{Y}$; a statement of this form should be taken to mean that X takes values in $\mathcal{X}$ and Y takes values in $\mathcal{Y}$.

Example* 1.4** ***(Binomial distribution). Consider the experiment considered in Example 1.2. Recall that an element ω of Ω is of the form $(x_1, \ldots, x_n)$ where each x_j is either 0 or 1. For an element $\omega \in \Omega$, define

$$X(\omega) = \sum_{j=1}^{n} x_j.$$

Then

$$\begin{aligned} \Pr(X = 0) &= \mathrm{P}((0, 0, \ldots, 0)) = (1 - \theta)^n, \\ \Pr(X = 1) &= \mathrm{P}((1, 0, \ldots, 0)) + \mathrm{P}((0, 1, 0, \ldots, 0)) + \cdots + \mathrm{P}((0, 0, \ldots, 0, 1)) \\ &= n\theta(1 - \theta)^{n-1}. \end{aligned}$$

It is straightforward to show that

$$\Pr(X = x) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}, \quad x = 0, 1, \ldots, n;$$

X is said to have a *binomial distribution* with parameters n and θ. □

Example* 1.5** ***(Uniform distribution on the unit cube). Let X denote a three-dimensional random vector with range $\mathcal{X} = (0, 1)^3$. For any subset of $A \in \mathcal{X}$, let

$$\Pr(X \in A) = \int\int\int_A dt_1\, dt_2\, dt_3.$$

Here the properties of the random vector X are defined without reference to any underlying experiment.

As discussed above, we may take the range of X to be $\mathbf{R}^3$. Then, for any subset $A \in \mathbf{R}^3$,

$$\Pr(X \in A) = \int\int\int_{A \cap (0,1)^3} dt_1\, dt_2\, dt_3.$$ □

Let X denote random variable on $\mathbf{R}^d$ with a given probability distribution. A *support* of the distribution, or, more simply, a support of X, is defined to be any set $\mathcal{X}_0 \subset \mathbf{R}^d$ such that

$$\Pr(X \in \mathcal{X}_0) = 1.$$

The *minimal support* of the distribution is the smallest closed set $\mathcal{X}_0 \subset \mathbf{R}^d$ such that

$$\Pr(X \in \mathcal{X}_0) = 1.$$

That is, the minimal support of X is a closed set $\mathcal{X}_0$ that is a support of X, and if $\mathcal{X}_1$ is another closed set that is a support of X, then $\mathcal{X}_0 \subset \mathcal{X}_1$.

The distribution of a real-valued random variable X is said to be *degenerate* if there exists a constant c such that

$$\Pr(X = c) = 1.$$

For a random vector X, with dimension greater than 1, the distribution of X is said to be degenerate if there exists a vector $a \neq 0$, with the same dimension as X, such that $a^T X$ is equal to a constant with probability 1. For example, a two-dimensional random vector $X = (X_1, X_2)$ has a degenerate distribution if, as in the case of a real-valued random variable, it is equal to a constant with probability 1. However, it also has a degenerate distribution if

$$\Pr(a_1 X_1 + a_2 X_2 = c) = 1$$

for some constants a_1, a_2, c. In this case, one of the components of X is redundant, in the sense that it can be expressed in terms of the other component (with probability 1).

Example* 1.6** ***(Polytomous random variable). Let X denote a random variable with range

$$\mathcal{X} = \{x_1, \ldots, x_m\}$$

where $x_1, \ldots, x_n$ are distinct elements of $\mathbf{R}$. Assume that $\Pr(X = x_j) > 0$ for each $j = 1, \ldots, m$. Any set containing $\mathcal{X}$ is a support of X; since $\mathcal{X}$ is closed in $\mathbf{R}$, it follows that the minimal support of X is simply $\mathcal{X}$. If $m = 1$ the distribution of X is degenerate; otherwise it is nondegenerate. □

Example* 1.7** ***(Uniform distribution on the unit cube). Let X denote the random variable defined in Example 1.5. Recall that for any $A \subset \mathbf{R}^3$,

$$\Pr(X \in A) = \int\int\int_{A \cap (0,1)^3} dt_1\, dt_2\, dt_3.$$

The minimal support of X is $[0, 1]^3$. □

Example* 1.8** ***(Degenerate random vector). Consider the experiment considered in Example 1.2 and used in Example 1.4 to define the binomial distribution. Recall that an element ω of Ω is of the form $(x_1, \ldots, x_n)$ where each x_j is either 0 or 1. Define Y to be the two-dimensional random vector given by

$$Y(\omega) = \left(\sum_{j=1}^n x_j, 2 \sum_{j=1}^n x_j^2 \right).$$

Then

$$\Pr((2, -1)^T Y = 0) = 1.$$

Hence, Y has a degenerate distribution. □

1.4 Distribution Functions

Consider a real-valued random variable X. The properties of X are described by its probability function P_X, which gives the probability that $X \in A$ for any set $A \subset \mathbf{R}$. However, it is also possible to specify the distribution of a random variable by considering $\Pr(X \in A)$ for a limited class of sets A; this approach has the advantage that the function giving such probabilities may be easier to use in computations. For instance, consider sets of the form $(-\infty, x]$, for $x \in \mathbf{R}$, so that $\mathrm{P}_X\{(-\infty, x]\}$ gives $\Pr(X \leq x)$. The *distribution function* of the distribution of X or, simply, the distribution function of X, is the function $F \equiv F_X : \mathbf{R} \to [0, 1]$ given by

$$F(x) = \Pr(X \leq x), \quad -\infty < x < \infty.$$

Example* 1.9 *(Uniform distribution). Suppose that X is a real-valued random variable such that

$$\Pr(X \in A) = \int_{A \cap (0,1)} dx, \quad A \subset \mathbf{R};$$

X is said to have a uniform distribution on $(0, 1)$.

The distribution function of this distribution is given by

$$F(x) = \Pr\{X \in (-\infty, x]\} = \int_{(-\infty, x] \cap (0,1)} dx = \begin{cases} 0 & \text{if } x \leq 0 \\ x & \text{if } 0 < x \leq 1. \\ 1 & \text{if } x > 1 \end{cases}$$

Figure 1.1 gives a plot of F. □

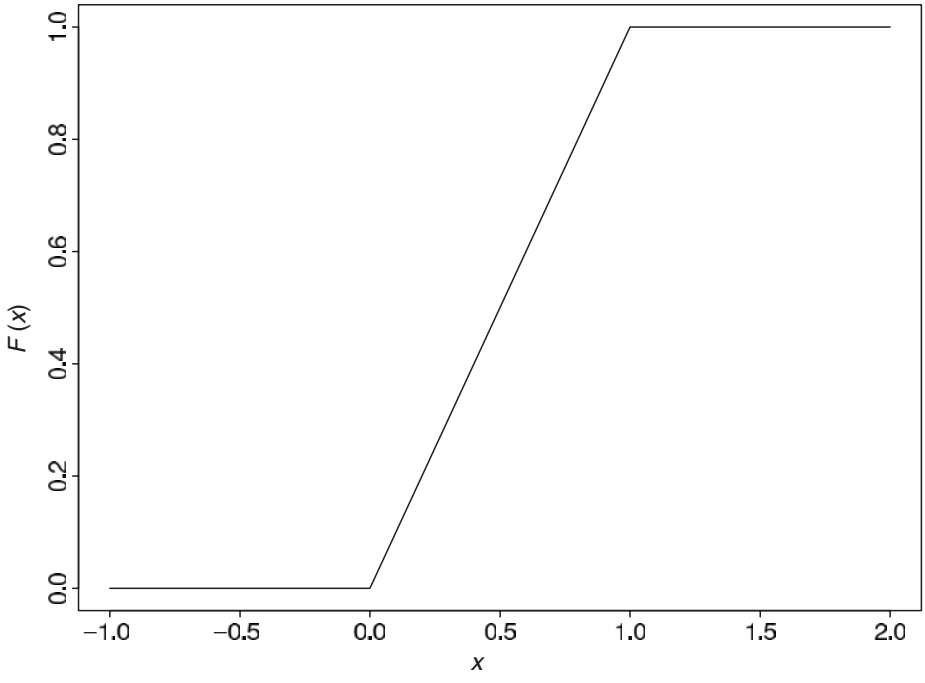

Figure 1.1. Distribution function in Example 1.9.

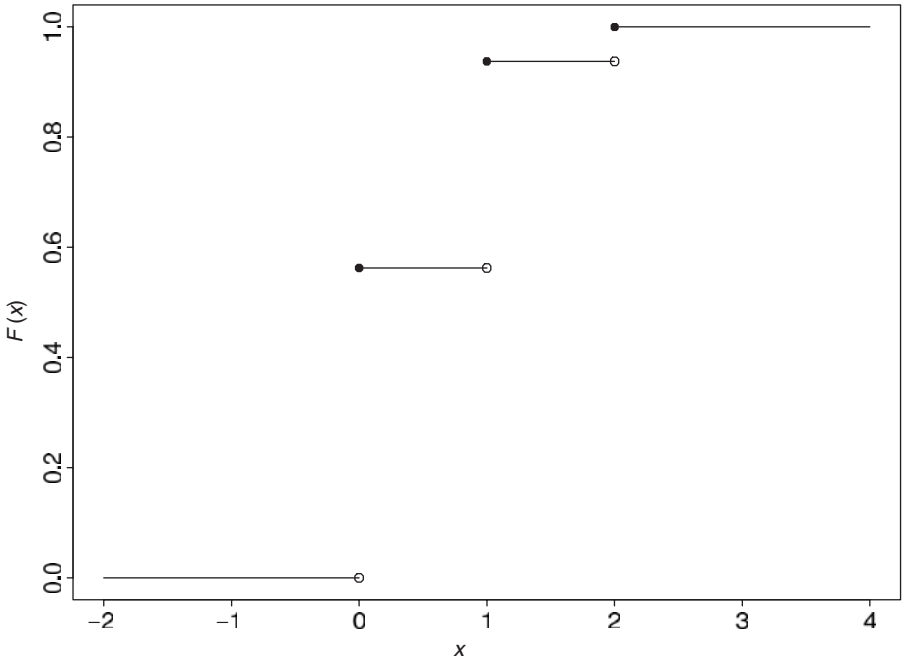

Figure 1.2. Distribution function in Example 1.10.

Note that when giving the form of a distribution function, it is convenient to only give the value of the function in the range of x for which $F(x)$ varies between 0 and 1. For instance, in the previous example, we might say that $F(x) = x$, $0 < x < 1$; in this case it is understood that $F(x) = 0$ for $x \leq 0$ and $F(x) = 1$ for $x \geq 1$.

Example* 1.10** ***(Binomial distribution). Let X denote a random variable with a binomial distribution with parameters n and θ, as described in Example 1.4. Then

$$\Pr(X = x) = \binom{n}{x} \theta^x (1-\theta)^{n-x}, \quad x = 0, 1, \ldots, n$$

and, hence, the distribution function of X is

$$F(x) = \sum_{j=0,1,\ldots;j \leq x} \binom{n}{j} \theta^j (1-\theta)^{n-j}.$$

Thus, F is a step function, with jumps at $0, 1, 2, \ldots, n$; Figure 1.2 gives a plot of F for the case $n = 2$, $\theta = 1/4$. □

Clearly, there are some basic properties which any distribution function F must possess. For instance, as noted above, F must take values in $[0, 1]$; also, F must be nondecreasing. The properties of a distribution function are summarized in the following theorem.

***Theorem* 1.2.** *A distribution function F of a distribution on* $\mathbf{R}$ *has the following properties:*

(DF1) $\lim_{x \to \infty} F(x) = 1$; $\lim_{x \to -\infty} F(x) = 0$

(DF2) If $x_1 < x_2$ *then* $F(x_1) \leq F(x_2)$

(DF3) $\lim_{h \to 0^+} F(x + h) = F(x)$

(DF4) $\lim_{h\to 0^+} F(x-h) \equiv F(x-) = F(x) - \Pr(X = x) = \Pr(X < x)$.

Proof. Let a_n, $n = 1, 2, \ldots$ denote any increasing sequence diverging to ∞ and let A_n denote the event that $X \leq a_n$. Then $\mathrm{P}_X(A_n) = F(a_n)$ and $A_1 \subset A_2 \subset \cdots$ with $\cup_{n=1}^{\infty} A_n$ equal to the event that $X < \infty$. It follows from (P4) that

$$\lim_{n\to\infty} F(a_n) = \Pr(X < \infty) = 1,$$

establishing the first part of (DF1); the second part follows in a similar manner.

To show (DF2), let A_1 denote the event that $X \leq x_1$ and A_2 denote the event that $x_1 < X \leq x_2$. Then A_1 and A_2 are disjoint with $F(x_1) = \mathrm{P}_X(A_1)$ and $F(x_2) = \mathrm{P}_X(A_1 \cup A_2) = \mathrm{P}_X(A_1) + \mathrm{P}_X(A_2)$, which establishes (DF2).

For (DF3) and (DF4), let a_n, $n = 1, 2, \ldots$, denote any decreasing sequence converging to 0, let A_n denote the event that $X \leq x + a_n$, let B_n denote the event that $X \leq x - a_n$, and let C_n denote the event that $x - a_n < X \leq x$. Then $A_1 \supset A_2 \supset \cdots$ and $\cap_{n=1}^{\infty} A_n$ is the event that $X \leq x$. Hence, by (P5),

$$\Pr(X \leq x) \equiv F(x) = \lim_{n\to\infty} F(x + a_n),$$

which establishes (DF3).

Finally, note that $F(x) = \mathrm{P}_X(B_n) + \mathrm{P}_X(C_n)$ and that $C_1 \supset C_2 \supset \cdots$ with $\cap_{n=1}^{\infty} C_n$ equal to the event that $X = x$. Hence,

$$F(x) = \lim_{n\to\infty} F(x - a_n) + \lim_{n\to\infty} \mathrm{P}_X(C_n) = F(x-) + \Pr(X = x),$$

yielding (DF4). ■

Thus, according to (DF2), a distribution function is nondecreasing and according to (DF3), a distribution is right-continuous.

A distribution function F gives the probability of sets of the form $(-\infty, x]$. The following result gives expressions for the probability of other types of intervals in terms of F; the proof is left as an exercise. As in Theorem 1.2, here we use the notation

$$F(x-) = \lim_{h\to 0^+} F(x - h).$$

***Corollary* 1.1.** *Let X denote a real-valued random variable with distribution function F. Then, for $x_1 < x_2$,*

(i) $\Pr(x_1 < X \leq x_2) = F(x_2) - F(x_1)$
(ii) $\Pr(x_1 \leq X \leq x_2) = F(x_2) - F(x_1-)$
(iii) $\Pr(x_1 \leq X < x_2) = F(x_2-) - F(x_1-)$
(iv) $\Pr(x_1 < X < x_2) = F(x_2-) - F(x_1)$

Any distribution function possesses properties (DF1)–(DF4). Furthermore, properties (DF1)–(DF3) characterize a distribution function in the sense that a function having those properties must be a distribution function of some random variable.

***Theorem* 1.3.** *If a function $F : \mathbf{R} \to [0, 1]$ has properties* (DF1)–(DF3)*, then F is the distribution function of some random variable.*