CHAPTER 1

Probability

1.1 INTRODUCTION

The theory of probability had its origin in gambling and games of chance. It owes much to the curiosity of gamblers who pestered their friends in the mathematical world with all sorts of questions. Unfortunately, this association with gambling contributed to very slow and sporadic growth of probability theory as a mathematical discipline. The mathematicians of the day took little or no interest in the development of any theory but looked only at the combinatorial reasoning involved in each problem.

The first attempt at some mathematical rigor is credited to Laplace. In his monumental work, *Theorie analytique des probabilités* (1812), Laplace gave the classical definition of the probability of an event that can occur only in a finite number of ways as the proportion of the number of favorable outcomes to the total number of all possible outcomes, provided that all the outcomes are *equally likely*. According to this definition, computation of the probability of events was reduced to combinatorial counting problems. Even in those days, this definition was found inadequate. In addition to being circular and restrictive, it did not answer the question of what probability is; it only gave a practical method of computing the probabilities of some simple events.

An extension of the classical definition of Laplace was used to evaluate the probabilities of sets of events with infinite outcomes. The notion of *equal likelihood* of certain events played a key role in this development. According to this extension, if Ω is some region with a well-defined measure (length, area, volume, etc.), the probability that a point chosen *at random* lies in a subregion A of Ω is the ratio measure(A)/measure(Ω). Many problems of geometric probability were solved using this extension. The trouble is that one can define *at random* in any way one pleases, and different definitions lead to different answers. For example, Joseph Bertrand, in his book *Calcul des probabilités* (Paris, 1889), cited a number of problems in geometric probability where the result depended on the method of solution. In Example 1.3.9 we discuss the famous Bertrand paradox and show that in reality there is nothing paradoxical about Bertrand's paradoxes; once we define *probability* *spaces* carefully, the paradox is resolved. Nevertheless, difficulties encountered in the field of geometric probability have been largely responsible for the slow growth of probability theory and its tardy acceptance by mathematicians as a mathematical discipline.

The mathematical theory of probability as we know it today is of comparatively recent origin. It was A. N. Kolmogorov who axiomatized probability in his fundamental work, *Foundations of the Theory of Probability* (Berlin), in 1933. According to this development, random events are represented by sets and probability is just a *normed measure* defined on these sets. This measure-theoretic development not only provided a logically consistent foundation for probability theory but also joined it to the mainstream of modern mathematics.

In this book we follow Kolmogorov's axiomatic development. In Section 1.2 we introduce the notion of a sample space. In Section 1.3 we state Kolmogorov's axioms of probability and study some simple consequences of these axioms. Section 1.4 is devoted to the computation of probability on finite sample spaces. Section 1.5 deals with conditional probability and Bayes rule, and Section 1.6 examines the independence of events.

1.2 SAMPLE SPACE

In most branches of knowledge, experiments are a way of life. In probability and statistics, too, we concern ourselves with special types of experiments. Consider the following examples.

Example 1. A coin is tossed. Assuming that the coin does not land on the side, there are two possible outcomes of the experiment: heads and tails. On any performance of this experiment, one does not know what the outcome will be. The coin can be tossed as many times as desired.

Example 2. A roulette wheel is a circular disk divided into 38 equal sectors numbered from 0 to 36 and 00. A ball is rolled on the edge of the wheel, and the wheel is rolled in the opposite direction. One bets on any of the 38 numbers or some combination of them. One can also bet on a color, red or black. If the ball lands in the sector numbered 32, say, anybody who bet on 32, or a combination including 32, wins; and so on. In this experiment, all possible outcomes are known in advance, namely 00, 0, 1, 2, . . . , 36, but on any performance of the experiment there is uncertainty as to what the outcome will be, provided, of course, that the wheel is not rigged in any manner. Clearly, the wheel can be rolled any number of times.

Example 3. A manufacturer produces 12-in rulers. The experiment consists in measuring as accurately as possible the length of a ruler produced by the manufacturer. Because of errors in the production process, one does not know what the true length of the ruler selected will be. It is clear, however, that the length will be, say, between 11 and 13 in., or, if one wants to be safe, between 6 and 18 in.

Example 4. The length of life of a light bulb produced by a certain manufacturer is recorded. In this case one does not know what the length of life will be for the light bulb selected, but clearly one is aware in advance that it will be some number between 0 and ∞ hours.

The experiments described above have certain common features. For each experiment, we know in advance all possible outcomes; that is, there are no surprises in store after any performance of the experiment. On any performance of the experiment, however, we do not know what the specific outcome will be; that is, there is uncertainty about the outcome on any performance of the experiment. Moreover, the experiment can be repeated under identical conditions. These features describe a random (or statistical) experiment.

Definition 1. A *random* (or *statistical*) *experiment* is an experiment in which:

(a) All outcomes of the experiment are known in advance.
(b) Any performance of the experiment results in an outcome that is not known in advance.
(c) The experiment can be repeated under identical conditions.

In probability theory we study this uncertainty of a random experiment. It is convenient to associate with each such experiment a set Ω, the set of all possible outcomes of the experiment. To engage in any meaningful discussion about the experiment, we associate with Ω a σ-field $\mathcal{S}$ of subsets of Ω. We recall that a σ-field is a nonempty class of subsets of Ω that is closed under the formation of countable unions and complements and contains the null set $\emptyset$.

Definition 2. The *sample space* of a statistical experiment is a pair $(\Omega, \mathcal{S})$, where

(a) Ω is the set of all possible outcomes of the experiment.
(b) $\mathcal{S}$ is a σ-field of subsets of Ω.

The elements of Ω are called *sample points*. Any set $A \in \mathcal{S}$ is known as an *event*. Clearly, A is a collection of sample points. We say that an event A happens if the outcome of the experiment corresponds to a point in A. Each one-point set is known as a *simple* or *elementary event*. If the set Ω contains only a finite number of points, we say that $(\Omega, \mathcal{S})$ is a *finite sample space*. If Ω contains at most a countable number of points, we call $(\Omega, \mathcal{S})$ a *discrete sample space*. If, however, Ω contains uncountably many points, we say that $(\Omega, \mathcal{S})$ is an *uncountable sample space*. In particular, if $\Omega = \mathcal{R}_k$ or some rectangle in $\mathcal{R}_k$, we call it a *continuous sample space*.

Remark 1. The choice of $\mathcal{S}$ is an important one, and some remarks are in order. If Ω contains at most a countable number of points, we can always take $\mathcal{S}$ to be the class of all subsets of Ω. This is certainly a σ-field. Each one-point set is a member of $\mathcal{S}$ and is the fundamental object of interest. Every subset of Ω is an event. If Ω

has uncountably many points, the class of all subsets of Ω is still a σ-field, but it is much too large a class of sets to be of interest. One of the most important examples of an uncountable sample space is the case in which $\Omega = \mathcal{R}$ or Ω is an interval in $\mathcal{R}$. In this case we would like all one-point subsets of Ω and all intervals (closed, open, or semiclosed) to be events. We use our knowledge of analysis to specify $\mathcal{S}$. We will not go into detail here except to recall that the class of all semiclosed intervals $(a, b]$ generates a class $\mathfrak{B}_1$ that is a σ-field on $\mathcal{R}$. This class contains all one-point sets and all intervals (finite or infinite). We take $\mathcal{S} = \mathfrak{B}_1$. Since we will be dealing mostly with the one-dimensional case, we write $\mathfrak{B}$ instead of $\mathfrak{B}_1$. There are many subsets of $\mathcal{R}$ that are not in $\mathfrak{B}_1$, but we do not demonstrate this fact here. We refer the reader to Halmos [39], Royden [94], or Kolmogorov and Fomin [52] for further details.

Example 5. Let us toss a coin. The set Ω is the set of symbols H and T, where H denotes head and T represents tail. Also, $\mathcal{S}$ is the class of all subsets of Ω, namely, $\{\{H\}, \{T\}, \{H, T\}, \emptyset\}$. If the coin is tossed two times, then

$$\Omega = \{(H, H), (H, T), (T, H), (T, T)\},$$

and

$$\begin{aligned}\mathcal{S} = \{&\emptyset, \{(H, H)\}, \{(H, T)\}, \{(T, H)\}, \{(T, T)\}, \{(H, H), (H, T)\}, \{(H, H), (T, H)\},\\ &\{(H, H), (T, T)\}, \{(H, T), (T, H)\}, \{(T, T), (T, H)\}, \{(T, T),\\ &(H, T)\}, \{(H, H), (H, T), (T, H)\}, \{(H, H), (H, T), (T, T)\},\\ &\{(H, H), (T, H), (T, T)\}, \{(H, T), (T, H), (T, T)\}, \Omega\},\end{aligned}$$

where the first element of a pair denotes the outcome of the first toss, and the second element, the outcome of the second toss. The event *at least one head* consists of sample points (H, H), (H, T), (T, H). The event *at most one head* is the collection of sample points (H, T), (T, H), (T, T).

Example 6. A die is rolled n times. The sample space is the pair $(\Omega, \mathcal{S})$, where Ω is the set of all n-tuples $(x_1, x_2, \dots, x_n)$, $x_i \in \{1, 2, 3, 4, 5, 6\}$, $i = 1, 2, \dots, n$, and $\mathcal{S}$ is the class of all subsets of Ω. Ω contains 6^n elementary events. The event A that 1 shows at least once is the set

$$\begin{aligned}A &= \{(x_1, x_2, \dots, x_n)\colon \text{at least one of } x_i\text{'s is } 1\}\\ &= \Omega - \{(x_1, x_2, \dots, x_n)\colon \text{none of the } x_i\text{'s is } 1\}\\ &= \Omega - \{(x_1, x_2, \dots, x_n)\colon x_i \in \{2, 3, 4, 5, 6\},\ i = 1, 2, \dots, n\}.\end{aligned}$$

Example 7. A coin is tossed until the first head appears. Then

$$\Omega = \{H, (T, H), (T, T, H), (T, T, T, H), \dots\},$$

and S is the class of all subsets of Ω. An equivalent way of writing Ω would be to look at the number of tosses required for the first head. Clearly, this number can take values $1, 2, 3, \ldots$, so that Ω is the set of all positive integers. Thus S is the class of all subsets of positive integers.

Example 8. Consider a pointer that is free to spin about the center of a circle. If the pointer is spun by an impulse, it will finally come to rest at some point. On the assumption that the mechanism is not rigged in any manner, each point on the circumference is a possible outcome of the experiment. The set Ω consists of all points $0 \le x < 2\pi r$, where r is the radius of the circle. Every one-point set $\{x\}$ is a simple event, namely, that the pointer will come to rest at x. The events of interest are those in which the pointer stops at a point belonging to a specified arc. Here S is taken to be the Borel σ-field of subsets of $[0, 2\pi r)$.

Example 9. A rod of length l is thrown onto a flat table, which is ruled with parallel lines at distance $2l$. The experiment consists in noting whether or not the rod intersects one of the ruled lines.

Let r denote the distance from the center of the rod to the nearest ruled line, and let θ be the angle that the axis of the rod makes with this line (Fig. 1). Every outcome of this experiment corresponds to a point (r, θ) in the plane. As Ω we take the set of all points (r, θ) in $\{(r, \theta): 0 \le r \le l, 0 \le \theta < \pi\}$. For S we take the Borel σ-field, $\mathfrak{B}_2$, of subsets of Ω, that is, the smallest σ-field generated by rectangles of the form

$$\{(x, y): a < x \le b, c < y \le d, 0 \le a < b \le l, 0 \le c < d < \pi\}.$$

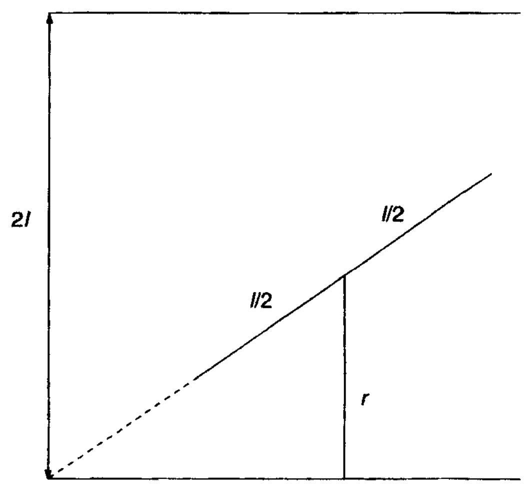

Fig. 1.

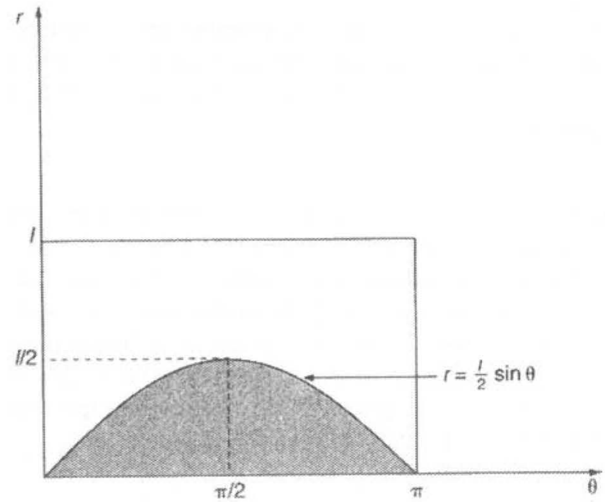

Fig. 2.

Clearly, the rod will intersect a ruled line if and only if the center of the rod lies in the area enclosed by the locus of the center of the rod (while one end touches the nearest line) and the nearest line (shaded area in Fig. 2).

Remark 2. From the discussion above it should be clear that in the discrete case there is really no problem. Every one-point set is also an event, and $\mathcal{S}$ is the class of all subsets of Ω. The problem, if there is any, arises only in regard to uncountable sample spaces. The reader has to remember only that in this case not all subsets of Ω are events. The case of most interest is the one in which $\Omega = \mathcal{R}_k$. In this case roughly all sets that have a well-defined volume (or area or length) are events. Not every set has the property in question, but sets that lack it are not easy to find and one does not encounter them in practice.

PROBLEMS 1.2

1. A club has five members, A, B, C, D, and E. It is required to select a chairman and a secretary. Assuming that one member cannot occupy both positions, write the sample space associated with these selections. What is the event that member A is an officeholder?

2. In each of the following experiments, what is the sample space?

(a) In a survey of families with three children, the genders of the children are recorded in increasing order of age.

(b) The experiment consists of selecting four items from a manufacturer's output and observing whether or not each item is defective.

(c) A given book is opened to any page, and the number of misprints is counted.

(d) Two cards are drawn from an ordinary deck of cards (i) with replacement, and (ii) without replacement.

3. Let A, B, C be three arbitrary events on a sample space $(\Omega, \mathcal{S})$. What is the event that only A occurs? What is the event that at least two of A, B, C occur? What is the event that both A and C, but not B, occur? What is the event that at most one of A, B, C occurs?

1.3 PROBABILITY AXIOMS

Let $(\Omega, \mathcal{S})$ be the sample space associated with a statistical experiment. In this section we define a probability set function and study some of its properties.

Definition 1. Let $(\Omega, \mathcal{S})$ be a sample space. A set function P defined on $\mathcal{S}$ is called a *probability measure* (or simply, *probability*) if it satisfies the following conditions:

(i) $P(A) \geq 0$ for all $A \in \mathcal{S}$.

(ii) $P(\Omega) = 1$.

(iii) Let $\{A_j\}$, $A_j \in \mathcal{S}$, $j = 1, 2, \dots$, be a disjoint sequence of sets; that is, $A_j \cap A_k = \emptyset$ for $j \neq k$, where $\emptyset$ is the null set. Then

$$P\left(\sum_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} P(A_j), \tag{1}$$

where we have used the notation $\sum_{j=1}^{\infty} A_j$ to denote union of disjoint sets A_j.

We call $P(A)$ the *probability of event* A. If there is no confusion, we will write PA instead of $P(A)$. Property (iii) is called *countable additivity*. That $P\emptyset = 0$ and P is also finitely additive follows from it.

Remark 1. If Ω is discrete and contains at most n $(< \infty)$ points, each single-point set $\{\omega_j\}$, $j = 1, 2, \dots, n$, is an elementary event, and it is sufficient to assign probability to each $\{\omega_j\}$. Then if $A \in \mathcal{S}$, where $\mathcal{S}$ is the class of all subsets of Ω, $PA = \sum_{\omega \in A} P\{\omega\}$. One such assignment is the *equally likely* assignment or the assignment of *uniform* probabilities. According to this assignment, $P\{\omega_j\} = 1/n$, $j = 1, 2, \dots, n$. Thus $PA = m/n$ if A contains m elementary events, $1 \leq m \leq n$.

Remark 2. If Ω is discrete and contains a countable number of points, one cannot make an equally likely assignment of probabilities. It suffices to make the assignment for each elementary event. If $A \in \mathcal{S}$, where $\mathcal{S}$ is the class of all subsets of Ω, define $PA = \sum_{\omega \in A} P\{\omega\}$.

Remark 3. If Ω contains uncountably many points, each one-point set is an elementary event, and again one cannot make an equally likely assignment of probabilities. Indeed, one cannot assign positive probability to each elementary event without violating the axiom $P\Omega = 1$. In this case one assigns probabilities to compound events consisting of intervals. For example, if $\Omega = [0, 1]$ and $\mathcal{S}$ is the Borel σ-field of all subsets of Ω, the assignment $P[I] =$ length of I, where I is a subinterval of Ω, defines a probability.

Definition 2. The triple $(\Omega, \mathcal{S}, P)$ is called a *probability space*.

Definition 3. Let $A \in \mathcal{S}$. We say that the *odds for A* are a to b if $PA = a/(a+b)$, and then the *odds against A* are b to a.

In many games of chance, probability is often stated in terms of odds against an event. Thus in horse racing a two-dollar bet on a horse to win with odds of 2 to 1 (against) pays approximately six dollars if the horse wins the race. In this case the probability of winning is $\frac{1}{3}$.

Example 1. Let us toss a coin. The sample space is $(\Omega, \mathcal{S})$, where $\Omega = \{\text{H}, \text{T}\}$ and $\mathcal{S}$ is the σ-field of all subsets of Ω. Let us define P on $\mathcal{S}$ as follows:

$$P\{\text{H}\} = \tfrac{1}{2} \quad \text{and} \quad P\{\text{T}\} = \tfrac{1}{2}.$$

Then P clearly defines a probability. Similarly, $P\{\text{H}\} = \frac{2}{3}$, $P\{\text{T}\} = \frac{1}{3}$, and $P\{\text{H}\} = 1$, $P\{\text{T}\} = 0$ are probabilities defined on $\mathcal{S}$. Indeed,

$$P\{\text{H}\} = p \quad \text{and} \quad P\{\text{T}\} = 1 - p \qquad (0 \le p \le 1)$$

defines a probability on $(\Omega, \mathcal{S})$.

Example 2. Let $\Omega = \{1, 2, 3, \ldots\}$ be the set of positive integers, and let $\mathcal{S}$ be the class of all subsets of Ω. Define P on $\mathcal{S}$ as follows:

$$P\{i\} = \frac{1}{2^i}, \qquad i = 1, 2, \ldots.$$

Then $\sum_{i=1}^{\infty} P\{i\} = 1$, and P defines a probability.

Example 3. Let $\Omega = (0, \infty)$ and $\mathcal{S} = \mathfrak{B}$, the Borel σ-field on Ω. Define P as follows: For each interval $I \subseteq \Omega$,

$$PI = \int_I e^{-x}\, dx.$$

Clearly, $PI \ge 0$, $P\Omega = 1$, and P is countably additive by properties of integrals.

Theorem 1. P is monotone and subtractive; that is, if $A, B \in \mathcal{S}$ and $A \subseteq B$, then $PA \leq PB$ and $P(B - A) = PB - PA$, where $B - A = B \cap A^c$, A^c being the complement of the event A.

Proof. If $A \subseteq B$, then

$$B = (A \cap B) + (B - A) = A + (B - A),$$

and it follows that $PB = PA + P(B - A)$.

Corollary. For all $A \in \mathcal{S}$, $0 \leq PA \leq 1$.

Remark 4. We wish to emphasize that if $PA = 0$ for some $A \in \mathcal{S}$, we call A an event with *zero probability* or a *null* event. However, it does not follow that $A = \emptyset$. Similarly, if $PB = 1$ for some $B \in \mathcal{S}$, we call B a *certain event*, but it does not follow that $B = \Omega$.

Theorem 2 (Addition Rule). If $A, B \in \mathcal{S}$, then

$$P(A \cup B) = PA + PB - P(A \cap B). \tag{2}$$

Proof. Clearly,

$$A \cup B = (A - B) + (B - A) + (A \cap B)$$

and

$$A = (A \cap B) + (A - B), B = (A \cap B) + (B - A).$$

The result follows by countable additivity of P.

Corollary 1. P is subadditive, that is, if $A, B \in \mathcal{S}$, then

$$P(A \cup B) \leq PA + PB. \tag{3}$$

Corollary 1 can be extended to an arbitrary number of events A_j,

$$P\left(\bigcup_j A_j\right) \leq \sum_j PA_j. \tag{4}$$

Corollary 2. If $B = A^c$, then A and B are disjoint and

$$PA = 1 - PA^c. \tag{5}$$

The following generalization of (2) is left as an exercise.

Theorem 3 (Principle of Inclusion–Exclusion). Let $A_1, A_2, \dots, A_n \in \mathcal{S}$. Then

$$P\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} PA_k - \sum_{k_1<k_2}^{n} P(A_{k_1} \cap A_{k_2}) + \sum_{k_1<k_2<k_3}^{n} P(A_{k_1} \cap A_{k_2} \cap A_{k_3}) + \cdots + (-1)^{n+1} P\left(\bigcap_{k=1}^{n} A_k\right). \tag{6}$$

Example 4. A die is rolled twice. Let all the elementary events in $\Omega = \{(i, j)\colon i, j = 1, 2, \dots, 6\}$ be assigned the same probability. Let A be the event that the first throw shows a number ≤ 2, and B be the event that the second throw shows at least 5. Then

$$\begin{aligned} A &= \{(i, j)\colon 1 \leq i \leq 2,\ j = 1, 2, \dots, 6\}, \\ B &= \{(i, j)\colon 5 \leq j \leq 6,\ i = 1, 2, \dots, 6\}, \\ A \cap B &= \{(1, 5), (1, 6), (2, 5), (2, 6)\}; \end{aligned}$$

and

$$\begin{aligned} P(A \cup B) &= PA + PB - P(A \cap B) \\ &= \tfrac{1}{3} + \tfrac{1}{3} - \tfrac{4}{36} = \tfrac{5}{9}. \end{aligned}$$

Example 5. A coin is tossed three times. Let us assign equal probability to each of the 2^3 elementary events in Ω. Let A be the event that at least one head shows up in three throws. Then

$$\begin{aligned} P(A) &= 1 - P(A^c) \\ &= 1 - P(\text{no heads}) \\ &= 1 - P(\text{TTT}) = \tfrac{7}{8}. \end{aligned}$$

We next derive two useful inequalities.

Theorem 4 (Bonferroni's Inequality). Given n (> 1) events $A_1, A_2, \dots, A_n$,

$$\sum_{i=1}^{n} PA_i - \sum_{i<j} P(A_i \cap A_j) \leq P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} PA_i. \tag{7}$$

Proof. In view of (4), it suffices to prove the left side of (7). The proof is by induction. The inequality on the left is true for $n = 2$ since

$$PA_1 + PA_2 - P(A_1 \cap A_2) = P(A_1 \cup A_2).$$

For $n = 3$,

$$P\left(\bigcup_{i=1}^{3} A_i\right) = \sum_{i=1}^{3} PA_i - \sum_{i<j} P(A_i \cap A_j) + P(A_1 \cap A_2 \cap A_3),$$

and the result holds. Assuming that (7) holds for $3 < m \leq n - 1$, we show that it also holds for $m + 1$:

$$\begin{aligned}
P\left(\bigcup_{i=1}^{m+1} A_i\right) &= P\left(\left(\bigcup_{i=1}^{m} A_i\right) \cup A_{m+1}\right) \\
&= P\left(\bigcup_{i=1}^{m} A_i\right) + PA_{m+1} - P\left(A_{m+1} \cap \left(\bigcup_{1}^{m} A_i\right)\right) \\
&\geq \sum_{i=1}^{m+1} PA_i - \sum_{i<j}^{m} P(A_i \cap A_j) - P\left(\bigcup_{i=1}^{m} (A_i \cap A_{m+1})\right) \\
&\geq \sum_{i=1}^{m+1} PA_i - \sum_{i<j}^{m} P(A_i \cap A_j) - \sum_{i=1}^{m} P(A_i \cap A_{m+1}) \\
&= \sum_{i=1}^{m+1} PA_i - \sum_{i<j}^{m+1} P(A_i \cap A_j).
\end{aligned}$$

Theorem 5 (Boole's Inequality). For any two events A and B,

$$P(A \cap B) \geq 1 - PA^c - PB^c. \tag{8}$$

Corollary 1. Let $\{A_j\}$, $j = 1, 2, \ldots$, be a countable sequence of events; then

$$P(\cap A_j) \geq 1 - \sum P(A_j^c). \tag{9}$$

Proof. Take

$$B = \bigcap_{j=2}^{\infty} A_j \quad \text{and} \quad A = A_1$$

in (8).

Corollary 2 (Implication Rule). If $A, B, C \in \mathcal{S}$ and A and B imply C, then

$$PC^c \leq PA^c + PB^c. \tag{10}$$

Let $\{A_n\}$ be a sequence of sets. The set of all points $\omega \in \Omega$ that belong to A_n for infinitely many values of n is known as the *limit superior* of the sequence and is denoted by

$$\limsup_{n\to\infty} A_n \quad \text{or} \quad \overline{\lim_{n\to\infty}}\, A_n.$$

The set of all points that belong to A_n for all but a finite number of values of n is known as the *limit inferior* of the sequence $\{A_n\}$ and is denoted by

$$\liminf_{n\to\infty} A_n \quad \text{or} \quad \underline{\lim_{n\to\infty}}\, A_n.$$

If

$$\underline{\lim_{n\to\infty}}\, A_n = \overline{\lim_{n\to\infty}}\, A_n,$$

we say that the limit exists and write $\lim_{n\to\infty} A_n$ for the common set and call it the *limit set*.

We have

$$\underline{\lim_{n\to\infty}}\, A_n = \bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} A_k \subseteq \bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k = \overline{\lim_{n\to\infty}}\, A_n.$$

If the sequence $\{A_n\}$ is such that $A_n \subseteq A_{n+1}$, for $n = 1, 2, \ldots$, it is called *nondecreasing*; if $A_n \supseteq A_{n+1}$, $n = 1, 2, \ldots$, it is called *nonincreasing*. If the sequence A_n is nondecreasing, we write $A_n \uparrow$; if A_n is nonincreasing, we write $A_n \downarrow$. Clearly, if $A_n \downarrow$ or $A_n \uparrow$, the limit exists and we have

$$\lim_n A_n = \bigcup_{n=1}^{\infty} A_n \qquad \text{if } A_n \uparrow$$

and

$$\lim_n A_n = \bigcap_{n=1}^{\infty} A_n \qquad \text{if } A_n \downarrow.$$

Theorem 6. Let $\{A_n\}$ be a nondecreasing sequence of events in $\mathcal{S}$; that is, $A_n \in \mathcal{S}$, $n = 1, 2, \ldots$, and

$$A_n \supseteq A_{n-1}, \qquad n = 2, 3, \ldots.$$

Then

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

Proof. Let

$$A = \bigcup_{j=1}^{\infty} A_j.$$

Then

$$A = A_n + \sum_{j=n}^{\infty}(A_{j+1} - A_j).$$

By countable additivity we have

$$PA = PA_n + \sum_{j=n}^{\infty} P(A_{j+1} - A_j).$$

and letting $n \to \infty$, we see that

$$PA = \lim_{n\to\infty} PA_n + \lim_{n\to\infty} \sum_{j=n}^{\infty} P(A_{j+1} - A_j).$$

The second term on the right tends to zero as $n \to \infty$ since the sum $\sum_{j=1}^{\infty} P(A_{j+1} - A_j) \leq 1$ and each summand is nonnegative. The result follows.

Corollary. Let $\{A_n\}$ be a nonincreasing sequence of events in $\mathcal{S}$. Then

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

Proof. Consider the nondecreasing sequence of events $\{A_n^c\}$. Then

$$\lim_{n\to\infty} A_n^c = \bigcup_{j=1}^{\infty} A_j^c = A^c.$$

It follows from Theorem 6 that

$$\lim_{n\to\infty} PA_n^c = P\left(\lim_{n\to\infty} A_n^c\right) = P\left(\bigcup_{j=1}^{\infty} A_n^c\right) = P(A^c).$$

In other words,

$$\lim_{n\to\infty}(1 - PA_n) = 1 - PA,$$

as asserted.

Remark 5. Theorem 6 and its corollary will be used quite frequently in subsequent chapters. Property (11) is called the *continuity of P from below*, and (12) is known as the *continuity of P from above*. Thus Theorem 6 and its corollary assure us that the set function P is continuous from above and below.

We conclude this section with some remarks concerning the use of the word *random* in this book. In probability theory *random* has essentially three meanings. First, in sampling from a finite population, a sample is said to be a *random sample* if at each draw all members available for selection have the same probability of being included. We discuss sampling from a finite population in Section 1.4. Second, we speak of a *random sample from a probability distribution*. This notion is formalized in Section 7.2. The third meaning arises in the context of geometric probability, where statements such as "a point is chosen randomly from the interval (a, b)" and "a point is picked randomly from a unit square" are frequently encountered. Once we have studied random variables and their distributions, problems involving geometric probabilities may be formulated in terms of problems involving independent uniformly distributed random variables, and these statements can be given appropriate interpretations.

Roughly speaking, these statements involve a certain assignment of probability. The word *random* expresses our desire to assign equal probability to sets of equal lengths, areas, or volumes. Let $\Omega \subseteq \mathcal{R}_n$ be a given set, and A be a subset of Ω. We are interested in the probability that a *randomly chosen point* in Ω falls in A. Here *randomly chosen* means that the point may be any point of Ω and that the probability of its falling in some subset A of Ω is proportional to the measure of A (independent of the location and shape of A). Assuming that both A and Ω have well-defined finite measures (length, area, volume, etc.), we define

$$PA = \frac{\text{measure}(A)}{\text{measure}(\Omega)}.$$

[In the language of measure theory we are assuming that Ω is a measurable subset of $\mathcal{R}_n$ that has a finite, positive Lebesgue measure. If A is any measurable set, $PA = \mu(A)/\mu(\Omega)$, where μ is the n-dimensional Lebesgue measure.] Thus, if a point is chosen at random from the interval (a, b), the probability that it lies in the interval $(c, d), a \le c < d \le b$, is $(d-c)/(b-a)$. Moreover, the probability that the randomly selected point lies in any interval of length $(d - c)$ is the same.

We present some examples.

Example 6. A point is picked "at random" from a unit square. Let $\Omega = \{(x, y): 0 \le x \le 1, 0 \le y \le 1\}$. It is clear that all rectangles and their unions must be in

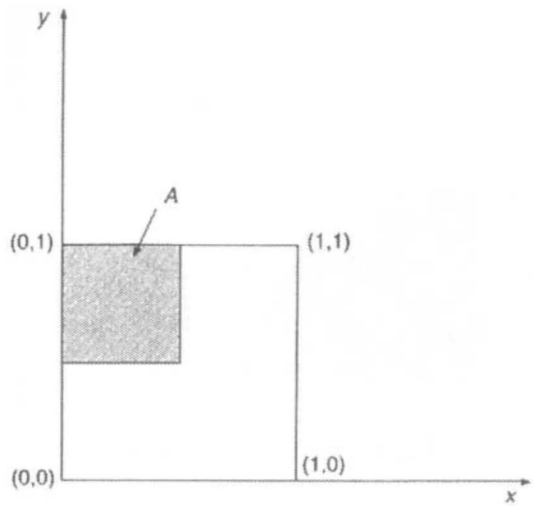

Fig. 1. $A = \{(x, y): 0 \le x \le \frac{1}{2}, \frac{1}{2} \le y \le 1\}$.

$\mathcal{S}$; so, too, should be all circles in the unit square, since the area of a circle is also well defined. Indeed, every set that has a well-defined area has to be in $\mathcal{S}$. We choose $\mathcal{S} = \mathfrak{B}_2$, the Borel σ-field generated by rectangles in Ω. As for the probability assignment, if $A \in \mathcal{S}$, we assign PA to A, where PA is the area of the set A. If $A = \{(x, y): 0 \le x \le \frac{1}{2}, \frac{1}{2} \le y \le 1\}$, then $PA = \frac{1}{4}$. If B is a circle with center $(\frac{1}{2}, \frac{1}{2})$ and radius $\frac{1}{2}$, then $PB = \pi(\frac{1}{2})^2 = \pi/4$. If C is the set of all points that are at most a unit distance from the origin, then $PC = \pi/4$ (see Figs. 1 to 3).

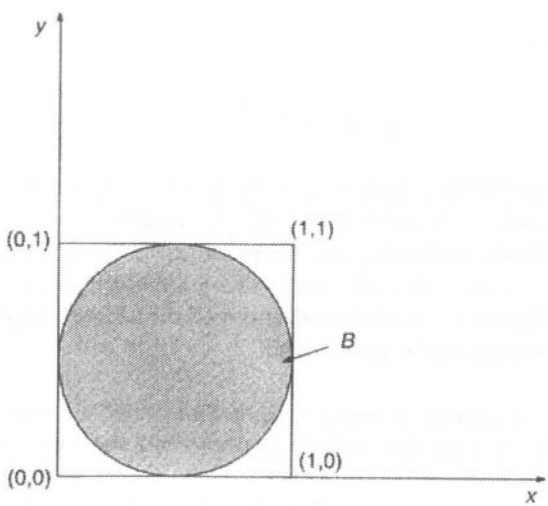

Fig. 2. $B = \{(x, y): (x - \frac{1}{2})^2 + (y - \frac{1}{2})^2 = 1\}$.

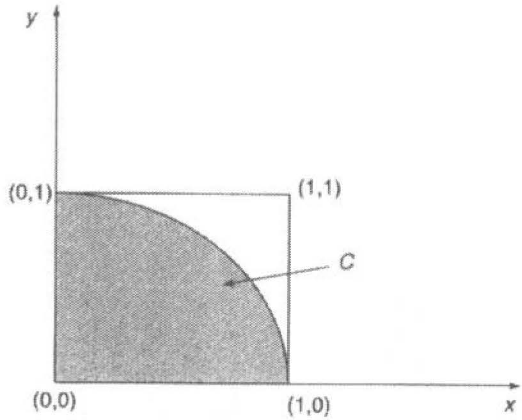

Fig. 3. $C = \{(x, y) : x^2 + y^2 \leq 1\}$.

Example 7 **(Buffon's Needle Problem).** We return to Example 1.2.9. A needle (rod) of length l is tossed at random on a plane that is ruled with a series of parallel lines a distance $2l$ apart. We wish to find the probability that the needle will intersect one of the lines. Denoting by r the distance from the center of the needle to the closest line and by θ the angle that the needle forms with this line, we see that a necessary and sufficient condition for the needle to intersect the line is that $r \leq (l/2) \sin\theta$. The needle will intersect the nearest line if and only if its center falls in the shaded region in Fig. 1.2.2. We assign probability to an event A as follows:

$$PA = \frac{\text{area of set } A}{l\pi}.$$

Thus the required probability is

$$\frac{1}{l\pi} \int_0^{\pi} \frac{l}{2} \sin\theta \, d\theta = \frac{1}{\pi}.$$

Here we have interpreted *at random* to mean that the position of the needle is characterized by a point (r, θ) which lies in the rectangle $0 \leq r \leq l$, $0 \leq \theta \leq \pi$. We have assumed that the probability that the point (r, θ) lies in any arbitrary subset of this rectangle is proportional to the area of this set. Roughly, this means that "all positions of the midpoint of the needle are assigned the same weight and all directions of the needle are assigned the same weight."

Example 8. An interval of length 1, say $(0, 1)$, is divided into three intervals by choosing two points at random. What is the probability that the three line segments form a triangle?

It is clear that a necessary and sufficient condition for the three segments to form a triangle is that the length of any one of the segments be less than the sum of the other two. Let x, y be the abscissas of the two points chosen at random. Then we must have either

$$0 < x < \tfrac{1}{2} < y < 1 \quad \text{and} \quad y - x < \tfrac{1}{2}$$

or

$$0 < y < \tfrac{1}{2} < x < 1 \quad \text{and} \quad x - y < \tfrac{1}{2}.$$

This is precisely the shaded area in Fig. 4. It follows that the required probability is $\frac{1}{4}$.

If it is specified in advance that the point x is chosen at random from $(0, \frac{1}{2})$, and the point y at random from $(\frac{1}{2}, 1)$, we must have

$$0 < x < \tfrac{1}{2}, \qquad \tfrac{1}{2} < y < 1,$$

and

$$y - x < x + 1 - y \quad \text{or} \quad 2(y - x) < 1.$$

In this case the area bounded by these lines is the shaded area in Fig. 5, and it follows that the required probability is $\frac{1}{2}$.

Note the difference in sample spaces in the two computations made above.

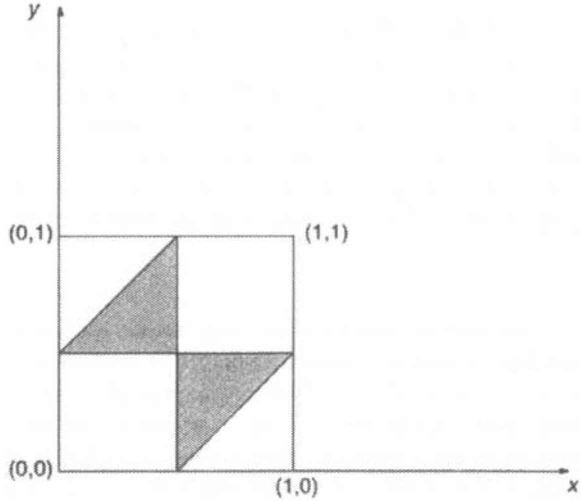

Fig. 4. $\{(x, y)\colon 0 < x < \frac{1}{2} < y < 1$, and $(y - x) < \frac{1}{2}$ or $0 < y < \frac{1}{2} < x < 1$, and $(x - y) < \frac{1}{2}\}$.

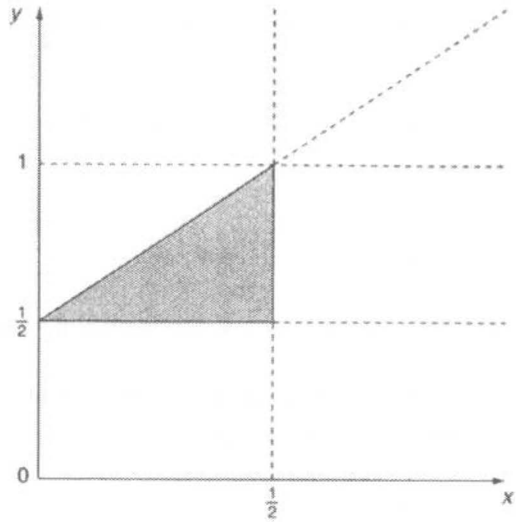

Fig. 5. $\{(x, y)\colon 0 < x < \frac{1}{2},\ \frac{1}{2} < y < 1 \text{ and } 2(y - x) < 1\}$.

Example 9 **(Bertrand's Paradox).** A chord is drawn at random in the unit circle. What is the probability that the chord is longer than the side of the equilateral triangle inscribed in the circle?

We present here three solutions to this problem, depending on how we interpret the phrase *at random*. The paradox is resolved once we define the probability spaces carefully.

SOLUTION 1. Since the length of a chord is uniquely determined by the position of its midpoint, choose a point C at random in the circle and draw a line through C and O, the center of the circle (Fig. 6). Draw the chord through C perpendicular to the line OC. If l_1 is the length of the chord with C as midpoint, $l_1 > \sqrt{3}$ if and only if C lies inside the circle with center O and radius $\frac{1}{2}$. Thus $PA = \pi(\frac{1}{2})^2/\pi = \frac{1}{4}$.

In this case Ω is the circle with center O and radius 1, and the event A is the concentric circle with center O and radius $\frac{1}{2}$. $\mathcal{S}$ is the usual Borel σ-field of subsets of Ω.

SOLUTION 2. Because of symmetry, we may fix one endpoint of the chord at some point P and then choose the other endpoint P_1 at random. Let the probability that P_1 lies on an arbitrary arc of the circle be proportional to the length of this arc. Now the inscribed equilateral triangle having P as one of its vertices divides the circumference into three equal parts. A chord drawn through P will be longer than the side of the triangle if and only if the other endpoint P_1 (Fig. 7) of the chord lies on that one-third of the circumference that is opposite P. It follows that the required probability is $\frac{1}{3}$. In this case $\Omega = [0, 2\pi]$, $\mathcal{S} = \mathfrak{B}_1 \cap \Omega$, and $A = [2\pi/3, 4\pi/3]$.

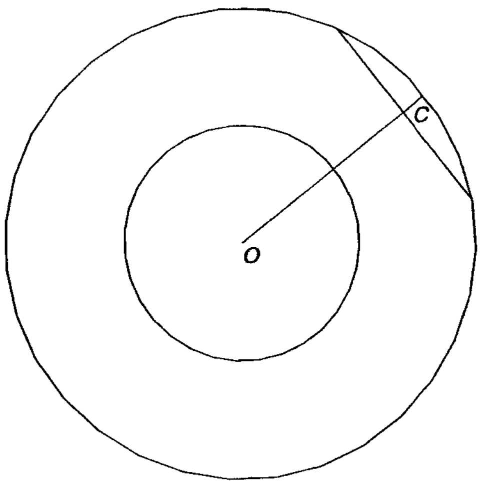

Fig. 6.

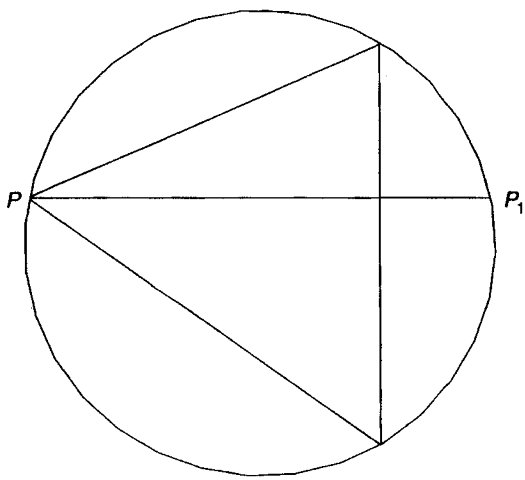

Fig. 7.

SOLUTION 3. Note that the length of a chord is determined uniquely by the distance of its midpoint from the center of the circle. Due to the symmetry of the circle, we assume that the midpoint of the chord lies on a fixed radius, OM, of the circle (Fig. 8). The probability that the midpoint M lies in a given segment of the radius through M is then proportional to the length of this segment. Clearly, the length of the chord will be longer than the side of the inscribed equilateral triangle if the length of OM is less than *radius*/2. It follows that the required probability is $\frac{1}{2}$.

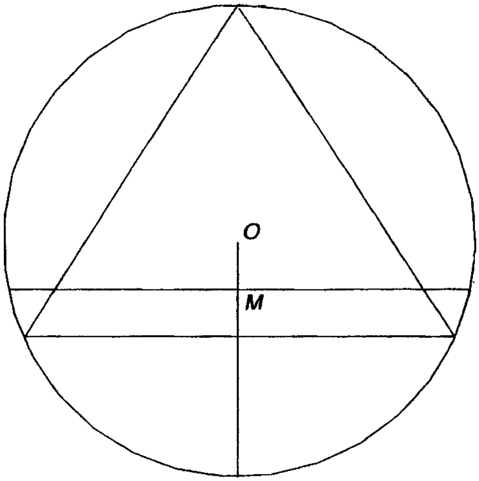

Fig. 8.

PROBLEMS 1.3

1. Let Ω be the set of all nonnegative integers and $\mathcal{S}$ the class of all subsets of Ω. In each of the following cases, does P define a probability on $(\Omega, \mathcal{S})$?

   (a) For $A \in \mathcal{S}$, let

   $$PA = \sum_{x \in A} \frac{e^{-\lambda}\lambda^x}{x!}, \qquad \lambda > 0.$$

   (b) For $A \in \mathcal{S}$, let

   $$PA = \sum_{x \in A} p(1-p)^x, \qquad 0 < p < 1.$$

   (c) For $A \in \mathcal{S}$, let $PA = 1$ if A has a finite number of elements, and $PA = 0$ otherwise.

2. Let $\Omega = \mathcal{R}$ and $\mathcal{S} = \mathfrak{B}$. In each of the following cases, does P define a probability on $(\Omega, \mathcal{S})$?

   (a) For each interval I, let

   $$PI = \int_I \frac{1}{\pi} \cdot \frac{1}{1+x^2}\, dx.$$

(b) For each interval I, let $PI = 1$ if I is an interval of finite length, and $PI = 0$ if I is an infinite interval.

(c) For each interval I, let $PI = 0$ if $I \subseteq (-\infty, 1)$ and $PI = \int_I (\frac{1}{2})\,dx$ if $I \subseteq [1, \infty]$. [If $I = I_1 + I_2$, where $I_1 \subseteq (-\infty, 1)$ and $I_2 \subseteq [1, \infty)$, then $PI = PI_2$.]

3. Let A and B be two events such that $B \supseteq A$. What is $P(A \cup B)$? What is $P(A \cap B)$? What is $P(A - B)$?

4. In Problem 1(a) and (b), let $A = \{\text{all integers} > 2\}$, $B = \{\text{all nonnegative integers} < 3\}$, and $C = \{\text{all integers } x,\ 3 < x < 6\}$. Find PA, PB, PC, $P(A \cap B)$, $P(A \cup B)$, $P(B \cup C)$, $P(A \cap C)$, and $P(B \cap C)$.

5. In Problem 2(a), let A be the event $A = \{x : x \geq 0\}$. Find PA. Also find $P\{x : x > 0\}$.

6. A box contains 1000 light bulbs. The probability that there is at least 1 defective bulb in the box is 0.1, and the probability that there are at least 2 defective bulbs is 0.05. Find the probability in each of the following cases:

(a) The box contains no defective bulbs.

(b) The box contains exactly 1 defective bulb.

(c) The box contains at most 1 defective bulb.

7. Two points are chosen at random on a line of unit length. Find the probability that each of the three line segments so formed will have a length $> \frac{1}{4}$.

8. Find the probability that the sum of two randomly chosen positive numbers (both ≤ 1) will not exceed 1 and that their product will be $\leq \frac{2}{9}$.

9. Prove Theorem 3.

10. Let $\{A_n\}$ be a sequence of events such that $A_n \to A$ as $n \to \infty$. Show that $PA_n \to PA$ as $n \to \infty$.

11. The base and altitude of a right triangle are obtained by picking points randomly from $[0, a]$ and $[0, b]$, respectively. Show that the probability that the area of the triangle so formed will be less than $ab/4$ is $(1 + \ln 2)/2$.

12. A point X is chosen at random on a line segment AB. (a) Show that the probability that the ratio of lengths AX/BX is smaller than a $(a > 0)$ is $a/(1 + a)$. (b) Show that the probability that the ratio of the length of the shorter segment to that of the larger segment is less than $\frac{1}{3}$ is $\frac{1}{2}$.

1.4 COMBINATORICS: PROBABILITY ON FINITE SAMPLE SPACES

In this section we restrict attention to sample spaces that have at most a finite number of points. Let $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ and $\mathcal{S}$ be the σ-field of all subsets of Ω. For any $A \in \mathcal{S}$,

$$PA = \sum_{\omega_j \in A} P\{\omega_j\}.$$

Definition 1. An assignment of probability is said to be *equally likely* (or *uniform*) if each elementary event in Ω is assigned the same probability. Thus, if Ω contains n points ω_j, $P\{\omega_j\} = 1/n$, $j = 1, 2, \ldots, n$.

With this assignment

$$PA = \frac{\text{number of elementary events in } A}{\text{total number of elementary events in } \Omega}. \tag{1}$$

Example 1. A coin is tossed twice. The sample space consists of four points. Under the uniform assignment, each of four elementary events is assigned probability $\frac{1}{4}$.

Example 2. Three dice are rolled. The sample space consists of 6^3 points. Each one-point set is assigned probability $1/6^3$.

In games of chance we usually deal with finite sample spaces where uniform probability is assigned to all simple events. The same is the case in sampling schemes. In such instances the computation of the probability of an event A reduces to a combinatorial counting problem. We therefore consider some rules of counting.

Rule 1. Given a collection of n_1 elements $a_{11}, a_{12}, \ldots, a_{1n_1}$, n_2 elements a_{21}, $a_{22}, \ldots, a_{2n_2}$, and so on, up to n_k elements $A_{k1}, a_{k2}, \ldots, a_{kn_k}$, it is possible to form $n_1 \cdot n_2 \cdot \cdots \cdot n_k$ ordered k-tuples $(a_{1j_1}, a_{2j_2}, \ldots, a_{kj_k})$ containing one element of each kind, $1 \le j_i \le n_i$, $i = 1, 2, \ldots, k$.

Example 3. Here r distinguishable balls are to be placed in n cells. This amounts to choosing one cell for each ball. The sample space consists of n^r r-tuples $(i_1, i_2, \ldots, i_r)$, where i_j is the cell number of the jth ball, $j = 1, 2, \ldots, r$ $(1 \le i_j \le n)$.

Consider r tossings with a coin. There are 2^r possible outcomes. The probability that no heads will show up in r throws is $(\frac{1}{2})^r$. Similarly, the probability that no 6 will turn up in r throws of a die is $(\frac{5}{6})^r$.

Rule 2 is concerned with *ordered samples*. Consider a set of n elements a_1, a_2, $\ldots, a_n$. Any ordered arrangement $(a_{i_1}, a_{i_2}, \ldots, a_{i_r})$ of r of these n symbols is called an *ordered sample* of size r. If elements are selected one by one, there are two possibilities:

1. *Sampling with replacement.* In this case repetitions are permitted, and we can draw samples of an arbitrary size. Clearly, there are n^r samples of size r.

2. *Sampling without replacement.* In this case an element once chosen is not replaced, so that there can be no repetitions. Clearly, the sample size cannot exceed n, the size of the population. There are $n(n-1)\cdots(n-r+1) = {}_nP_r$, say, possible samples of size r. Clearly, ${}_nP_r = 0$ for integers $r > n$. If $r = n$, then ${}_nP_r = n!$.

Rule 2. If ordered samples of size r are drawn from a population of n elements, there are n^r different samples with replacement and ${}_nP_r$ samples without replacement.

Corollary. The number of permutations of n objects is $n!$.

Remark 1. We frequently use the term *random sample* in this book to describe the equal assignment of probability to all possible samples in sampling from a finite population. Thus, when we speak of a random sample of size r from a population of n elements, it means that in sampling with replacement, each of n^r samples has the same probability $1/n^r$ or that in sampling without replacement, each of ${}_nP_r$ samples is assigned probability $1/{}_nP_r$.

Example 4. Consider a set of n elements. A sample of size r is drawn at random with replacement. Then the probability that no element appears more than once is clearly ${}_nP_r/n^r$.

Thus, if n balls are to be randomly placed in n cells, the probability that each cell will be occupied is $n!/n^n$.

Example 5. Consider a class of r students. The birthdays of these r students form a sample of size r from the 365 days in the year. Then the probability that all r birthdays are different is ${}_{365}P_r/(365)^r$. One can show that this probability is $< \frac{1}{2}$ if $r = 23$.

The following table gives the values of $q_r = {}_{365}P_r/(365)^r$ for some selected values of r.

r	20	23	25	30	35	60
q_r	0.589	0.493	0.431	0.294	0.186	0.006

Next suppose that each of the r students is asked for his or her birth date in order, with the instruction that as soon as a student hears his or her birth date the student is to raise a hand. Let us compute the probability that a hand is first raised when the kth ($k = 1, 2, \ldots, r$) student is asked his or her birth date. Let p_k be the probability that the procedure terminates at the kth student. Then

$$p_1 = 1 - \left(\frac{364}{365}\right)^{r-1}$$

and

$$p_k = \frac{{}_{365}P_{k-1}}{(365)^{k-1}}\left(1 - \frac{k-1}{365}\right)^{r-k+1}\left[1 - \left(\frac{365-k}{365-k+1}\right)^{r-k}\right], \quad k = 2, 3, \ldots, r.$$

Example 6. Let Ω be the set of all permutations of n objects. Let A_i be the set of all permutations that leave the ith object unchanged. Then the set $\bigcup_{i=1}^{n} A_i$ is the set of permutations with at least one fixed point. Clearly,

$$PA_i = \frac{(n-1)!}{n!}, \qquad i = 1, 2, \ldots, n,$$

$$P(A_i \cap A_j) = \frac{(n-2)!}{n!}, \qquad i < j; \quad i, j = 1, 2, \ldots, n, \text{ etc.}$$

By Theorem 1.3.3 we have

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \left(1 - \frac{1}{2!} + \frac{1}{3!} - \cdots \pm \frac{1}{n!}\right).$$

As an application, consider an absentminded secretary who places n letters in n envelopes at random. Then the probability that he or she will misplace every letter is

$$1 - \left(1 - \frac{1}{2!} + \frac{1}{3!} - \cdots \pm \frac{1}{n!}\right).$$

It is easy to see that this last probability $\longrightarrow e^{-1} = 0.3679$ as $n \to \infty$.

Rule 3. There are $\binom{n}{r}$ different subpopulations of size $r \le n$ from a population of n elements, where

$$\binom{n}{r} = \frac{n!}{r!\,(n-r)!}. \tag{2}$$

Example 7. Consider the random distribution of r balls in n cells. Let A_k be the event that a specified cell has exactly k balls, $k = 0, 1, 2, \ldots, r$; k balls can be chosen in $\binom{r}{k}$ ways. We place k balls in the specified cell and distribute the remaining $r - k$ balls in the $n - 1$ cells in $(n-1)^{r-k}$ ways. Thus

$$PA_k = \binom{r}{k}\frac{(n-1)^{r-k}}{n^r} = \binom{r}{k}\left(\frac{1}{n}\right)^k\left(1 - \frac{1}{n}\right)^{r-k}.$$

Example 8. There are $\binom{52}{13} = 635{,}013{,}559{,}600$ different hands at bridge and $\binom{52}{5} = 2{,}598{,}960$ hands at poker.

The probability that all 13 cards in a bridge hand have different face values is $4^{13}/\binom{52}{13}$.

The probability that a hand at poker contains five different face values is $\binom{13}{5}4^5/\binom{52}{5}$.

Rule 4. Consider a population of n elements. The number of ways in which the population can be partitioned into k subpopulations of sizes $r_1, r_2, \ldots, r_k$, respectively, $r_1 + r_2 + \cdots + r_k = n$, $0 \le r_i \le n$, is given by

$$\binom{n}{r_1, r_2, \ldots, r_k} = \frac{n!}{r_1!\, r_2! \cdots r_k!}. \tag{3}$$

The numbers defined in (3) are known as *multinomial coefficients*.

Proof. For the proof of Rule 4, one uses Rule 3 repeatedly. Note that

$$\binom{n}{r_1, r_2, \ldots, r_k} = \binom{n}{r_1}\binom{n-r_1}{r_2}\cdots\binom{n-r_1-\cdots-r_{k-2}}{r_{k-1}}. \tag{4}$$

Example 9. In a game of bridge the probability that a hand of 13 cards contains 2 spades, 7 hearts, 3 diamonds, and 1 club is

$$\frac{\binom{13}{2}\binom{13}{7}\binom{13}{3}\binom{13}{1}}{\binom{52}{13}}.$$

Example 10. An urn contains 5 red, 3 green, 2 blue, and 4 white balls. A sample of size 8 is selected at random without replacement. The probability that the sample contains 2 red, 2 green, 1 blue, and 3 white balls is

$$\frac{\binom{5}{2}\binom{3}{2}\binom{2}{1}\binom{4}{3}}{\binom{14}{8}}.$$

PROBLEMS 1.4

1. How many different words can be formed by permuting letters of the word *Mississippi*? How many of these start with the letters *Mi*?

2. An urn contains R red and W white marbles. Marbles are drawn from the urn one after another without replacement. Let A_k be the event that a red marble is drawn for the first time on the kth draw. Show that

$$PA_k = \frac{R}{R+W-k+1} \prod_{j=1}^{k-1} \left(1 - \frac{R}{R+W-j+1}\right).$$

Let p be the proportion of red marbles in the urn before the first draw. Show that $PA_k \to p(1-p)^{k-1}$ as $R+W \to \infty$. Is this to be expected?

3. In a population of N elements, R are red and $W = N - R$ are white. A group of n elements is selected at random. Find the probability that the group so chosen will contain exactly r red elements.

4. Each permutation of the digits 1, 2, 3, 4, 5, 6 determines a six-digit number. If the numbers corresponding to all possible permutations are listed in increasing order of magnitude, find the 319th number on this list.

5. The numbers $1, 2, \ldots, n$ are arranged in random order. Find the probability that the digits $1, 2, \ldots, k$ $(k < n)$ appear as neighbors in that order.

6. A pinball table has seven holes through which a ball can drop. Five balls are played. Assuming that at each play a ball is equally likely to go down any one of the seven holes, find the probability that more than one ball goes down at least one of the holes.

7. If $2n$ boys are divided into two equal subgroups, find the probability that the two tallest boys will be (a) in different subgroups, and (b) in the same subgroup.

8. In a movie theater that can accommodate $n+k$ people, n people are seated. What is the probability that $r \le n$ given seats are occupied?

9. Waiting in line for a Saturday morning movie show are $2n$ children. Tickets are priced at a quarter each. Find the probability that nobody will have to wait for change if before a ticket is sold to the first customer, the cashier has $2k$ $(k < n)$ quarters. Assume that it is equally likely that each ticket is paid for with a quarter or a half-dollar coin.

10. Each box of a certain brand of breakfast cereal contains a small charm, with k distinct charms forming a set. Assuming that the chance of drawing any particular charm is equal to that of drawing any other charm, show that the probability of finding at least one complete set of charms in a random purchase of $N \ge k$ boxes equals

$$1 - \binom{k}{1}\left(\frac{k-1}{k}\right)^N + \binom{k}{2}\left(\frac{k-2}{k}\right)^N - \binom{k}{3}\left(\frac{k-3}{k}\right)^N$$
$$+ \cdots + (-1)^{k-1}\binom{k}{k-1}\left(\frac{1}{k}\right)^N. \qquad [\textit{Hint:}\ \text{Use (1.3.7).}]$$

11. Prove Rules 1 through 4.

12. In a five-card poker game, find the probability that a hand will have:

(a) A royal flush (ace, king, queen, jack, and 10 of the same suit).

(b) A straight flush (five cards in a sequence, all of the same suit; ace is high but A, 2, 3, 4, 5 is also a sequence), excluding a royal flush.

(c) Four of a kind (four cards of the same face value).

(d) A full house (three cards of the same face value x and two cards of the same face value y).

(e) A flush (five cards of the same suit, excluding cards in a sequence).

(f) A straight (five cards in a sequence).

(g) Three of a kind (three cards of the same face value and two cards of different face values).

(h) Two pairs.

(i) A single pair.

13. (a) A married couple and four of their friends enter a row of seats in a concert hall. What is the probability that the wife will sit next to her husband if all possible seating arrangements are equally likely?

(b) In part (a), suppose that the six people go to a restaurant after the concert and sit at a round table. What is the probability that the wife will sit next to her husband?

14. Consider a town with N people. A person sends two letters to two separate people, each of whom is asked to repeat the procedure. Thus for each letter received, two letters are sent out to separate persons chosen at random (irrespective of what happened in the past). What is the probability that in the first n stages the person who started the chain letter game will not receive a letter?

15. Consider a town with N people. A person tells a rumor to a second person, who in turn repeats it to a third person, and so on. Suppose that at each stage the recipient of the rumor is chosen at random from the remaining $N - 1$ people. What is the probability that the rumor will be repeated n times:

(a) Without being repeated to any person?

(b) Without being repeated to the originator?

16. There were four accidents in a town during a seven–day period. Would you be surprised if all four occurred on the same day? If each of the four occurred on a different day?

17. Whereas Rules 1 and 2 of counting deal with ordered samples with or without replacement, Rule 3 concerns unordered sampling without replacement. The most difficult rule of counting deals with unordered with replacement sampling. Show that there are $\binom{n+r-1}{r}$ possible unordered samples of size r from a population of n elements when sampled with replacement.

1.5 CONDITIONAL PROBABILITY AND BAYES THEOREM

So far, we have computed probabilities of events on the assumption that no information was available about the experiment other than the sample space. Sometimes, however, it is known that an event H has happened. How do we use this information in making a statement concerning the outcome of another event A? Consider the following examples.

Example 1. Let urn 1 contain one white and two black balls, and urn 2, one black and two white balls. A fair coin is tossed. If a head turns up, a ball is drawn at random from urn 1; otherwise, from urn 2. Let E be the event that the ball drawn is black. The sample space is $\Omega = \{\mathrm{H}b_{11}, \mathrm{H}b_{12}, \mathrm{H}w_{11}, \mathrm{T}b_{21}, \mathrm{T}w_{21}, \mathrm{T}w_{22}\}$, where H denotes head, T denotes tail, b_{ij} denotes jth black ball in ith urn, $i = 1, 2$, and so on. Then

$$PE = P\{\mathrm{H}b_{11}, \mathrm{H}b_{12}, \mathrm{T}b_{21}\} = \tfrac{3}{6} = \tfrac{1}{2}.$$

If, however, it is known that the coin showed a head, the ball could not have been drawn from urn 2. Thus, the probability of E, conditional on information H, is $\frac{2}{3}$. Note that this probability equals the ratio $P\{\text{head and ball drawn black}\}/P\{\text{head}\}$.

Example 2. Let us toss two fair coins. Then the sample space of the experiment is $\Omega = \{\mathrm{HH}, \mathrm{HT}, \mathrm{TH}, \mathrm{TT}\}$. Let event $A = \{\text{both coins show same face}\}$ and $B = \{\text{at least one coin shows H}\}$. Then $PA \doteq \frac{2}{4}$. If B is known to have happened, this information assures that TT cannot happen, and $P\{A \text{ conditional on the information that } B \text{ has happened}\} = \frac{1}{3} = \frac{1}{4}/\frac{3}{4} = P(A \cap B)/PB$.

Definition 1. Let $(\Omega, \mathcal{S}, P)$ be a probability space, and let $H \in \mathcal{S}$ with $PH > 0$. For an arbitrary $A \in \mathcal{S}$ we shall write

$$P\{A \mid H\} = \frac{P(A \cap H)}{PH} \tag{1}$$

and call the quantity so defined the *conditional probability* of A, given H. Conditional probability remains undefined when $PH = 0$.

Theorem 1. Let $(\Omega, \mathcal{S}, P)$ be a probability space, and let $H \in \mathcal{S}$ with $PH > 0$. Then $(\Omega, \mathcal{S}, P_H)$, where $P_H(A) = P\{A \mid H\}$ for all $A \in \mathcal{S}$, is a probability space.

Proof. Clearly, $P_H(A) = P\{A \mid H\} \geq 0$ for all $A \in \mathcal{S}$. Also, $P_H(\Omega) = P(\Omega \cap H)/PH = 1$. If $A_1, A_2, \ldots$ is a disjoint sequence of sets in $\mathcal{S}$, then

$$P_H\left(\sum_{i=1}^{\infty} A_i\right) = P\left\{\sum_{i=1}^{\infty} A_i \mid H\right\} = \frac{P\left\{\left(\sum_1^{\infty} A_i\right) \cap H\right\}}{PH}$$

$$= \frac{\sum_{i=1}^{\infty} P(A_i \cap H)}{PH} = \sum_{i=1}^{\infty} P_H(A_i).$$

Remark 1. What we have done is to consider a new sample space consisting of the basic set H and the σ-field $\mathcal{S}_H = \mathcal{S} \cap H$, of subsets $A \cap H$, $A \in \mathcal{S}$, of H. On this space we have defined a set function P_H by multiplying the probability of each event by $(PH)^{-1}$. Indeed, $(H, \mathcal{S}_H, P_H)$ is a probability space.

Let A and B be two events with $PA > 0$, $PB > 0$. Then it follows from (1) that

$$(2) \qquad P(A \cap B) = PA \cdot P\{B \mid A\}, \quad \text{and} \quad P(A \cap B) = PB \cdot P\{A \mid B\}.$$

Equations (2) may be generalized to any number of events. Let $A_1, A_2, \ldots, A_n \in \mathcal{S}$, $n \geq 2$, and assume that $P(\bigcap_{j=1}^{n-1} A_i) > 0$. Since

$$A_1 \supset (A_1 \cap A_2) \supset (A_1 \cap A_2 \cap A_3) \supset \cdots \supset \left(\bigcap_{j=1}^{n-2} A_j\right) \supset \left(\bigcap_{j=1}^{n-1} A_j\right),$$

we see that

$$PA_1 > 0, \quad P(A_1 \cap A_2) > 0, \quad \ldots, \quad P\left(\bigcap_{j=1}^{n-2} A_j\right) > 0.$$

It follows that $P\{A_k \mid \cap_{j=1}^{k-1} A_j\}$ are well defined for $k = 2, 3, \ldots, n$.

Theorem 2 (Multiplication Rule). Let $(\Omega, \mathcal{S}, P)$ be a probability space and $A_1, A_2, \ldots, A_n \in \mathcal{S}$, with $P(\cap_{j=1}^{n-1} A_j) > 0$. Then

$$(3) \quad P\left(\bigcap_{j=1}^{n} A_j\right) = P(A_1)P\{A_2 \mid A_1\}P\{A_3 \mid A_1 \cap A_2\} \cdots P\left\{A_n \,\middle|\, \bigcap_{j=1}^{n-1} A_j\right\}.$$

Proof. The proof is simple.

Let us suppose that $\{H_j\}$ is a countable collection of events in $\mathcal{S}$ such that $H_j \cap H_k = \emptyset$, $j \neq k$, and $\sum_{j=1}^{\infty} H_j = \Omega$. Suppose that $PH_j > 0$ for all j. Then

$$(4) \qquad PB = \sum_{j=1}^{\infty} P(H_j)P\{B \mid H_j\} \qquad \text{for all } B \in \mathcal{S}.$$

For the proof we note that

$$B = \sum_{j=1}^{\infty} (B \cap H_j),$$

and the result follows. Equation (4) is called the *total probability rule*.

Example 3. Consider a hand of five cards in a game of poker. If the cards are dealt at random, there are $\binom{52}{5}$ possible hands of five cards each. Let $A = \{$at least 3 cards of spades$\}$, $B = \{$all 5 cards of spades$\}$. Then

$$P(A \cap B) = P\{\text{all 5 cards of spades}\} = \frac{\binom{13}{5}}{\binom{52}{5}}$$

and

$$P\{B \mid A\} = \frac{P(A \cap B)}{PA} = \frac{\binom{13}{5}\Big/\binom{52}{5}}{\left[\binom{13}{3}\binom{39}{2} + \binom{13}{4}\binom{39}{1} + \binom{13}{5}\right]\Big/\binom{52}{5}}.$$

Example 4. Urn 1 contains one white and two black marbles, urn 2 contains one black and two white marbles, and urn 3 contains three black and three white marbles. A die is rolled. If a 1, 2, or 3 shows up, urn 1 is selected; if a 4 shows up, urn 2 is selected; and if a 5 or 6 shows up, urn 3 is selected. A marble is then drawn at random from the urn selected. Let A be the event that the marble drawn is white. If U, V, W, respectively, denote the events that the urn selected is 1, 2, 3, then

$$A = (A \cap U) + (A \cap V) + (A \cap W),$$
$$P(A \cap U) = P(U) \cdot P\{A \mid U\} = \tfrac{3}{6} \cdot \tfrac{1}{3},$$
$$P(A \cap V) = P(V) \cdot P\{A \mid V\} = \tfrac{1}{6} \cdot \tfrac{2}{3},$$
$$P(A \cap W) = P(W) \cdot P\{A \mid W\} = \tfrac{2}{6} \cdot \tfrac{3}{6}.$$

It follows that

$$PA = \tfrac{1}{6} + \tfrac{1}{9} + \tfrac{1}{6} = \tfrac{4}{9}.$$

A simple consequence of the total probability rule is the Bayes rule, which we now prove.

Theorem 3 (Bayes Rule). Let $\{H_n\}$ be a disjoint sequence of events such that $PH_n > 0, n = 1, 2, \ldots$, and $\sum_{n=1}^{\infty} H_n = \Omega$. Let $B \in \mathcal{S}$ with $PB > 0$. Then

(5) $$P\{H_j \mid B\} = \frac{P(H_j)P\{B \mid H_j\}}{\sum_{i=1}^{\infty} P(H_i)P\{B \mid H_i\}}, \qquad j = 1, 2, \ldots.$$

Proof. From (2)

$$P\{B \cap H_j\} = P(B)P\{H_j \mid B\} = PH_jP\{B \mid H_j\},$$

and it follows that

$$P\{H_j \mid B\} = \frac{PH_jP\{B \mid H_j\}}{PB}.$$

The result now follows on using (4).

Remark 2. Suppose that $H_1, H_2, \ldots$ are all the "causes" that lead to the outcome of a random experiment. Let H_j be the set of outcomes corresponding to the jth cause. Assume that the probabilities PH_j, $j = 1, 2, \ldots$, called the *prior probabilities*, can be assigned. Now suppose that the experiment results in an event B of positive probability. This information leads to a reassessment of the prior probabilities. The conditional probabilities $P\{H_j \mid B\}$ are called the *posterior probabilities*. Formula (5) can be interpreted as a rule giving the probability that observed event B was due to cause or hypothesis H_j.

Example 5. In Example 4, let us compute the conditional probability $P\{V \mid A\}$. We have

$$P\{V \mid A\} = \frac{PVP\{A \mid V\}}{PUP\{A \mid U\} + PVP\{A \mid V\} + PWP\{A \mid W\}}$$
$$= \frac{\frac{1}{6} \cdot \frac{2}{3}}{\frac{3}{6} \cdot \frac{1}{3} + \frac{1}{6} \cdot \frac{2}{3} + \frac{2}{6} \cdot \frac{3}{6}} = \frac{\frac{1}{9}}{\frac{4}{9}} = \frac{1}{4}.$$

PROBLEMS 1.5

1. Let A and B be two events such that $PA = p_1 > 0$, $PB = p_2 > 0$, and $p_1 + p_2 > 1$. Show that $P\{B \mid A\} \geq 1 - [(1 - p_2)/p_1]$.
2. Two digits are chosen at random without replacement from the set of integers $\{1, 2, 3, 4, 5, 6, 7, 8\}$.
   (a) Find the probability that both digits are greater than 5.
   (b) Show that the probability that the sum of the digits will be equal to 5 is the same as the probability that their sum will exceed 13.
3. The probability of a family chosen at random having exactly k children is αp^k, $0 < p < 1$. Suppose that the probability that any child has blue eyes is b,

$0 < b < 1$, independently of others. What is the probability that a family chosen at random has exactly r ($r \geq 0$) children with blue eyes?

4. In Problem 3, let us write

$$\begin{aligned} p_k &= \text{probability of a randomly chosen family having exactly } k \text{ children} \\ &= \alpha p^k, \qquad k = 1, 2, \dots, \\ p_0 &= 1 - \frac{\alpha p}{1-p}. \end{aligned}$$

Suppose that all gender distributions of k children are equally likely. Find the probability that a family has exactly r boys, $r \geq 1$. Find the conditional probability that a family has at least two boys, given that it has at least one boy.

5. Each of $(N + 1)$ identical urns marked $0, 1, 2, \dots, N$ contains N balls. The kth urn contains k black and $N - k$ white balls, $k = 0, 1, 2, \dots, N$. An urn is chosen at random, and n random drawings are made from it, the ball drawn always being replaced. If all the n draws result in black balls, find the probability that the $(n + 1)$th draw will also produce a black ball. How does this probability behave as $N \to \infty$?

6. Each of n urns contains four white and six black balls, while another urn contains five white and five black balls. An urn is chosen at random from the $(n+1)$ urns, and two balls are drawn from it, both being black. The probability that five white and three black balls remain in the chosen urn is $\frac{1}{7}$. Find n.

7. In answering a question on a multiple-choice test, a candidate either knows the answer with probability p ($0 \leq p < 1$) or does not know the answer with probability $1 - p$. If he knows the answer, he puts down the correct answer with probability 0.99, whereas if he guesses, the probability of his putting down the correct result is $1/k$ (k choices to the answer). Find the conditional probability that the candidate knew the answer to a question, given that he has made the correct answer. Show that this probability tends to 1 as $k \to \infty$.

8. An urn contains five white and four black balls. Four balls are transferred to a second urn. A ball is then drawn from this urn, and it happens to be black. Find the probability of drawing a white ball from among the remaining three.

9. Prove Theorem 2.

10. An urn contains r red and g green marbles. A marble is drawn at random and its color noted. Then the marble drawn, together with $c > 0$ marbles of the same color, are returned to the urn. Suppose that n such draws are made from the urn. Find the probability of selecting a red marble at any draw.

11. Consider a bicyclist who leaves a point P (see Fig. 1), choosing one of the roads PR_1, PR_2, PR_3 at random. At each subsequent crossroad she again chooses a road at random.

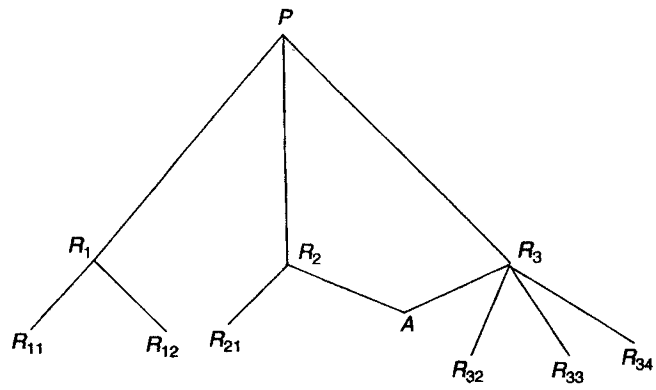

Fig. 1. Map for Problem 11.

(a) What is the probability that she will arrive at point A?

(b) What is the conditional probability that she will arrive at A via road PR_3?

12. Five percent of patients suffering from a certain disease are selected to undergo a new treatment that is believed to increase the recovery rate from 30 percent to 50 percent. A person is randomly selected from these patients after the completion of the treatment and is found to have recovered. What is the probability that the patient received the new treatment?

13. Four roads lead away from the county jail. A prisoner has escaped from the jail and selects a road at random. If road I is selected, the probability of escaping is $\frac{1}{8}$; if road II is selected, the probability of success is $\frac{1}{6}$; if road III is selected, the probability of escaping is $\frac{1}{4}$; and if road IV is selected, the probability of success is $\frac{9}{10}$.

(a) What is the probability that the prisoner will succeed in escaping?

(b) If the prisoner succeeds, what is the probability that the prisoner escaped by using road IV? By using road I?

14. A diagnostic test for a certain disease is 95 percent accurate, in that if a person has the disease, it will detect it with a probability of 0.95, and if a person does not have the disease, it will give a negative result with a probability of 0.95. Suppose that only 0.5 percent of the population has the disease in question. A person is chosen at random from this population. The test indicates that this person has the disease. What is the (conditional) probability that he or she does have the disease?

1.6 INDEPENDENCE OF EVENTS

Let $(\Omega, \mathcal{S}, P)$ be a probability space, and let $A, B \in \mathcal{S}$, with $PB > 0$. By the multiplication rule we have

$$P(A \cap B) = P(B)P\{A \mid B\}.$$

In many experiments the information provided by B does not affect the probability of event A; that is, $P\{A \mid B\} = P\{A\}$.

Example 1. Let two fair coins be tossed, and let $A = \{\text{head on the second throw}\}$, $B = \{\text{head on the first throw}\}$. Then

$$P(A) = P\{\text{HH, TH}\} = \tfrac{1}{2}, \qquad P(B) = \{\text{HH, HT}\} = \tfrac{1}{2},$$

and

$$P\{A \mid B\} = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{4}}{\frac{1}{2}} = \tfrac{1}{2} = P(A).$$

Thus

$$P(A \cap B) = P(A)P(B).$$

In the following, we write $A \cap B = AB$.

Definition 1. Two events, A and B, are said to be *independent* if and only if

$$P(AB) = P(A)P(B). \tag{1}$$

Note that we have not placed any restriction on $P(A)$ or $P(B)$. Thus conditional probability is not defined when $P(A)$ or $P(B) = 0$, but independence is. Clearly, if $P(A) = 0$, then A is independent of every $E \in \mathcal{S}$. Also, any event $A \in \mathcal{S}$ is independent of $\emptyset$ and Ω.

Theorem 1. If A and B are independent events, then

$$P\{A \mid B\} = P(A) \qquad \text{if } P(B) > 0$$

and

$$P\{B \mid A\} = P(B) \qquad \text{if } P(A) > 0.$$

Theorem 2. If A and B are independent, so are A and B^c, A^c and B, and A^c and B^c.

Proof.

$$\begin{aligned} P(A^cB) &= P(B - (A \cap B)) \\ &= P(B) - P(A \cap B) \qquad \text{since } B \supseteq (A \cap B) \\ &= P(B)\,[1 - P(A)] \\ &= P(A^c)\,P(B). \end{aligned}$$

Similarly, one proves that A^c and B^c, and A and B^c, are independent.

We wish to emphasize that independence of events is not to be confused with disjoint or mutually exclusive events. If two events, each with nonzero probability, are mutually exclusive, they are obviously dependent since the occurrence of one will automatically preclude the occurrence of the other. Similarly, if A and B are independent and $PA > 0$, $PB > 0$, then A and B cannot be mutually exclusive.

Example 2. A card is chosen at random from a deck of 52 cards. Let A be the event that the card is an ace, and B, the event that it is a club. Then

$$P(A) = \tfrac{4}{52} = \tfrac{1}{13}, \quad P(B) = \tfrac{13}{52} = \tfrac{1}{4},$$

and

$$P(AB) = P\{\text{ace of clubs}\} = \tfrac{1}{52},$$

so that A and B are independent.

Example 3. Consider families with two children, and assume that all four possible distributions of gender: BB, BG, GB, GG, where B stands for boy and G for girl, are equally likely. Let E be the event that a randomly chosen family has at most one girl, and F, the event that the family has children of both genders. Then

$$P(E) = \tfrac{3}{4}, \quad P(F) = \tfrac{1}{2}, \quad \text{and} \quad P(EF) = \tfrac{1}{2},$$

so that E and F are not independent.

Now consider families with three children. Assuming that each of the eight possible gender distributions is equally likely, we have

$$P(E) = \tfrac{4}{8}, \quad P(F) = \tfrac{6}{8}, \quad \text{and} \quad P(EF) = \tfrac{3}{8},$$

so that E and F are independent.

An obvious extension of the concept of independence between two events A and B to a given collection $\mathfrak{U}$ of events is to require that any two distinct events in $\mathfrak{U}$ be independent.

Definition 2. Let $\mathfrak{U}$ be a family of events from $\mathcal{S}$. We say that the events $\mathfrak{U}$ are *pairwise independent* if and only if for every pair of distinct events $A, B \in \mathfrak{U}$,

$$P(AB) = PA\, PB.$$

A much stronger and more useful concept is mutual or complete independence.

Definition 3. A family of events $\mathfrak{U}$ is said to be a *mutually* or *completely independent* family if and only if for every finite subcollection $\{A_{i_1}, A_{i_2}, \ldots, A_{i_k}\}$ of $\mathfrak{U}$, the following relation holds:

$$(2) \qquad P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = \prod_{j=1}^{k} PA_{i_j}.$$

In what follows we omit the adjective *mutual* or *complete* and speak of independent events. It is clear from Definition 3 that to check the independence of n events $A_1, A_2, \ldots, A_n \in \mathcal{S}$, we must check the following $2^n - n - 1$ relations:

$$\begin{aligned}
P(A_iA_j) &= PA_iPA_j, & i \neq j;\ i, j = 1, 2, \ldots, n,\\
P(A_iA_jA_k) &= PA_iPA_jPA_k, & i \neq j \neq k;\ i, j, k = 1, 2, \ldots, n,\\
&\ \vdots\\
P(A_1A_2\cdots A_n) &= PA_1PA_2\cdots PA_n.
\end{aligned}$$

The first of these requirements is pairwise independence. Independence therefore implies pairwise independence, but not conversely.

Example 4 (Wong [119]). Take four identical marbles. On the first, write symbols $A_1A_2A_3$. On each of the other three, write A_1, A_2, A_3, respectively. Put the four marbles in an urn and draw one at random. Let E_i denote the event that the symbol A_i appears on the drawn marble. Then

$$P(E_1) = P(E_2) = P(E_3) = \tfrac{1}{2},$$

$$P(E_1E_2) = P(E_2E_3) = P(E_1E_3) = \tfrac{1}{4},$$

and

$$(3) \qquad P(E_1E_2E_3) = \tfrac{1}{4}.$$

It follows that although events E_1, E_2, E_3, are not independent, they are pairwise independent.

Example 5 (Kac [46], pp. 22–23). In this example $P(E_1E_2E_3) = P(E_1) \times P(E_2)P(E_3)$, but E_1, E_2, E_3 are not pairwise independent and hence not independent. Let $\Omega = \{1, 2, 3, 4\}$, and let p_i be the probability assigned to $\{i\}$, $i = 1, 2, 3, 4$. Let $p_1 = \sqrt{2}/2 - \frac{1}{4}$, $p_2 = \frac{1}{4}$, $p_3 = \frac{3}{4} - \sqrt{2}/2$, $p_4 = \frac{1}{4}$. Let $E_1 = \{1, 3\}$, $E_2 = \{2, 3\}$, $E_3 = \{3, 4\}$. Then

$$\begin{aligned}
P(E_1E_2E_3) = P\{3\} = \frac{3}{4} - \frac{\sqrt{2}}{2} &= \frac{1}{2}\left(1 - \frac{\sqrt{2}}{2}\right)\left(1 - \frac{\sqrt{2}}{2}\right)\\
&= (p_1 + p_3)(p_2 + p_3)(p_3 + p_4)\\
&= P(E_1)P(E_2)P(E_3).
\end{aligned}$$

But $P(E_1E_2) = \frac{3}{4} - \sqrt{2}/2 \neq PE_1PE_2$, and it follows that E_1, E_2, E_3 are not independent.

Example 6. A die is rolled repeatedly until a 6 turns up. We will show that event A, that "a 6 will eventually show up," is certain to occur. Let A_k be the event that a 6 will show up for the first time on the kth throw. Let $A = \sum_{k=1}^{\infty} A_k$. Then

$$PA_k = \frac{1}{6}\left(\frac{5}{6}\right)^{k-1}, \qquad k = 1, 2, \ldots,$$

and

$$PA = \frac{1}{6}\sum_{k=1}^{\infty}\left(\frac{5}{6}\right)^{k-1} = \frac{1}{6}\frac{1}{1-\frac{5}{6}} = 1.$$

Alternatively, we can use the corollary to Theorem 1.3.6. Let B_n be the event that a 6 does not show up on the first n trials. Clearly, $B_{n+1} \subseteq B_n$, and we have $A^c = \bigcap_{n=1}^{\infty} B_n$. Thus

$$1 - PA = PA^c = P\left(\bigcap_{n=1}^{\infty} B_n\right) = \lim_{n\to\infty} P(B_n) = \lim_{n\to\infty}\left(\frac{5}{6}\right)^n = 0.$$

Example 7. A slip of paper is given to person A, who marks it with either a plus or minus sign; the probability of her writing a plus sign is $\frac{1}{3}$. A passes the slip to B, who may either leave it alone or change the sign before passing it to C. Next, C passes the slip to D after perhaps changing the sign; finally, D passes it to a referee after perhaps changing the sign. The referee sees a plus sign on the slip. It is known that B, C, and D each change the sign with probability $\frac{2}{3}$. We shall compute the probability that A originally wrote a plus.

Let N be the event that A wrote a plus sign, and M, the event that she wrote a minus sign. Let E be the event that the referee saw a plus sign on the slip. We have

$$P\{N \mid E\} = \frac{P(N)P\{E \mid N\}}{P(M)P\{E \mid M\} + P(N)P\{E \mid N\}}.$$

Now

$$\begin{aligned} P\{E \mid N\} &= P\{\text{the plus sign was either not changed or changed exactly twice}\} \\ &= \left(\frac{1}{3}\right)^3 + 3\left(\frac{2}{3}\right)^2 + \left(\frac{1}{3}\right) \end{aligned}$$

and

$$\begin{aligned} P\{E \mid M\} &= P\{\text{the minus sign was changed either once or three times}\} \\ &= 3\left(\frac{2}{3}\right)\left(\frac{1}{3}\right)^2 + \left(\frac{2}{3}\right)^3. \end{aligned}$$

It follows that

$$P\{N \mid E\} = \frac{(\frac{1}{3})[(\frac{1}{3})^3 + 3(\frac{2}{3})^2(\frac{1}{3})]}{(\frac{1}{3})[(\frac{1}{3})^3 + 3(\frac{2}{3})^2(\frac{1}{3})] + (\frac{2}{3})[3(\frac{2}{3})(\frac{1}{3})^2 + (\frac{2}{3})^3]}$$
$$= \frac{\frac{13}{18}}{\frac{41}{81}} = \frac{13}{41}.$$

PROBLEMS 1.6

1. A biased coin is tossed until a head appears for the first time. Let p be the probability of a head, $0 < p < 1$. What is the probability that the number of tosses required is odd? Even?

2. Let A and B be two independent events defined on some probability space, and let $PA = \frac{1}{3}$, $PB = \frac{3}{4}$. Find (a) $P(A \cup B)$, (b) $P\{A \mid A \cup B\}$, and (c) $P\{B \mid A \cup B\}$.

3. Let A_1, A_2, A_3 be three independent events. Show that A_1^c, A_2^c, and A_3^c are independent.

4. A biased coin with probability p, $0 < p < 1$, of success (heads) is tossed until for the first time, the same result occurs three times in succession (that is, three heads or three tails in succession). Find the probability that the game will end at the seventh throw.

5. A box contains 20 black and 30 green balls. One ball at a time is drawn at random, its color is noted, and the ball is then replaced in the box for the next draw.
   (a) Find the probability that the first green ball is drawn on the fourth draw.
   (b) Find the probability that the third and fourth green balls are drawn on the sixth and ninth draws, respectively.
   (c) Let N be the trial at which the fifth green ball is drawn. Find the probability that the fifth green ball is drawn on the nth draw. (Note that N take values $5, 6, 7, \ldots$.)

6. An urn contains four red and four black balls. A sample of two balls is drawn at random. If both balls drawn are of the same color, these balls are set aside and a new sample is drawn. If the two balls drawn are of different colors, they are returned to the urn and another sample is drawn. Assume that the draws are independent and that the same sampling plan is pursued at each stage until all balls are drawn.
   (a) Find the probability that at least n samples are drawn before two balls of the same color appear.
   (b) Find the probability that after the first two samples are drawn, four balls are left, two black and two red.

7. Let A, B, and C be three boxes with three, four, and five cells, respectively. There are three yellow balls numbered 1 to 3, four green balls numbered 1 to 4,

and five red balls numbered 1 to 5. The yellow balls are placed at random in box A, the green in B, and the red in C, with no cell receiving more than one ball. Find the probability that only one of the boxes will show no matches.

8. A pond contains red and golden fish. There are 3000 red and 7000 golden fish, of which 200 and 500, respectively, are tagged. Find the probability that a random sample of 100 red and 200 golden fish will show 15 and 20 tagged fish, respectively.

9. Let $(\Omega, \mathcal{S}, P)$ be a probability space. Let $A, B, C \in \mathcal{S}$ with PB and $PC > 0$. If B and C are independent, show that

$$P\{A \mid B\} = P\{A \mid B \cap C\}PC + P\{A \mid B \cap C^c\}PC^c.$$

Conversely, if this relation holds, $P\{A \mid BC\} \neq P\{A \mid B\}$, and $PA > 0$, then B and C are independent. (Strait [110])

10. Show that the converse of Theorem 2 also holds. Thus A and B are independent if, and only if, A and B^c are independent; and so on.

11. A lot of five identical batteries is life tested. The probability assignment is assumed to be

$$P(A) = \int_A \frac{1}{\lambda} e^{-x/\lambda}\, dx$$

for any event $A \subseteq [0, \infty)$, where $\lambda > 0$ is a known constant. Thus the probability that a battery fails after time t is given by

$$P(t, \infty) = \int_t^\infty \frac{1}{\lambda} e^{-x/\lambda}\, dx, \qquad t \geq 0.$$

If the times to failure of the batteries are independent, what is the probability that at least one battery will be operating after t_0 hours?

12. On $\Omega = (a, b)$, $-\infty < a < b < \infty$, each subinterval is assigned a probability proportional to the length of the interval. Find a necessary and sufficient condition for two events to be independent.

13. A game of craps is played with a pair of fair dice as follows. A player rolls the dice. If a sum of 7 or 11 shows up, the player wins; if a sum of 2, 3, or 12 shows up, the player loses. Otherwise, the player continues to roll the pair of dice until the sum is either 7 or the first number rolled. In the former case the player loses, and in the latter the player wins.

(a) Find the probability that the player wins on the nth roll.

(b) Find the probability that the player wins the game.

(c) What is the probability that the game ends on (i) the first roll, (ii) the second roll, and (iii) the third roll?

CHAPTER 2

Random Variables and Their Probability Distributions

2.1 INTRODUCTION

In Chapter 1 we dealt essentially with random experiments that can be described by finite sample spaces. We studied the assignment and computation of probabilities of events. In practice, one observes a function defined on the space of outcomes. Thus, if a coin is tossed n times, one is not interested in knowing which of the 2^n n-tuples in the sample space has occurred. Rather, one would like to know the number of heads in n tosses. In games of chance, one is interested in the net gain or loss of a certain player. Actually, in Chapter 1 we were concerned with such functions without defining the term *random variable*. Here we study the notion of a random variable and examine some of its properties.

In Section 2.2 we define a random variable, and in Section 2.3 we study the notion of probability distribution of a random variable. Section 2.4 deals with some special types of random variables, and in Section 2.5 we consider functions of a random variable and their induced distributions. The fundamental difference between a random variable and a real-valued function of a real variable is the associated notion of a probability distribution. Nevertheless, our knowledge of advanced calculus or real analysis is the basic tool in the study of random variables and their probability distributions.

2.2 RANDOM VARIABLES

In Chapter 1 we studied properties of a set function P defined on a sample space $(\Omega, \mathcal{S})$. Since P is a set function, it is not very easy to handle; we cannot perform arithmetic or algebraic operations on sets. Moreover, in practice one frequently observes some function of elementary events. When a coin is tossed repeatedly, which replication resulted in heads is not of much interest. Rather, one is interested in the number of heads, and consequently, the number of tails, that appear in, say, n tossings of the coin. It is therefore desirable to introduce a point function on the sample space. We can then use our knowledge of calculus or real analysis to study properties of P.

Definition 1. Let $(\Omega, \mathcal{S})$ be a sample space. A finite, single-valued function that maps Ω into $\mathcal{R}$ is called a *random variable* (RV) if the inverse images under X of all Borel sets in $\mathcal{R}$ are events, that is, if

$$X^{-1}(B) = \{\omega\colon X(\omega) \in B\} \in \mathcal{S} \qquad \text{for all } B \in \mathfrak{B}. \tag{1}$$

To verify whether a real-valued function on $(\Omega, \mathcal{S})$ is an RV, it is not necessary to check that (1) holds for all Borel sets $B \in \mathfrak{B}$. It suffices to verify (1) for any class $\mathfrak{A}$ of subsets of $\mathcal{R}$ that generates $\mathfrak{B}$. By taking $\mathfrak{A}$ to be the class of semiclosed intervals $(-\infty, x]$, $x \in \mathcal{R}$, we get the following result.

Theorem 1. X is an RV if and only if for each $x \in \mathcal{R}$,

$$\{\omega\colon X(\omega) \le x\} = \{X \le x\} \in \mathcal{S}. \tag{2}$$

Remark 1. Note that the notion of probability does not enter into the definition of an RV.

Remark 2. If X is an RV, the sets $\{X = x\}$, $\{a < X \le b\}$, $\{X < x\}$, $\{a \le X < b\}$, $\{a < X < b\}$, $\{a \le X \le b\}$ are all events. Moreover, we could have used any of these intervals to define an RV. For example, we could have used the following equivalent definition: X is an RV if and only if

$$\{\omega\colon X(\omega) < x\} \in \mathcal{S} \qquad \text{for all } x \in \mathcal{R}. \tag{3}$$

We have

$$\{X < x\} = \bigcup_{n=1}^{\infty} \left(X \le x - \frac{1}{n}\right) \tag{4}$$

and

$$\{X \le x\} = \bigcap_{n=1}^{\infty} \left(X < x + \frac{1}{n}\right). \tag{5}$$

Remark 3. In practice, (1) or (2) is a technical condition in the definition of an RV which the reader may ignore and think of RVs simply as real-valued functions defined on Ω. It should be emphasized, though, that there do exist subsets of $\mathcal{R}$ that do not belong to $\mathfrak{B}$, and hence there exist real-valued functions defined on Ω that are not RVs, but the reader will not encounter them in practical applications.

Example 1. For any set $A \subseteq \Omega$, define

$$I_A(\omega) = \begin{cases} 0, & \omega \notin A, \\ 1, & \omega \in A. \end{cases}$$

$I_A(\omega)$ is called the *indicator function* of set A. I_A is an RV if and only if $A \in \mathcal{S}$.

Example 2. Let $\Omega = \{\text{H, T}\}$, and $\mathcal{S}$ be the class of all subsets of Ω. Define X by $X(\text{H}) = 1$, $X(\text{T}) = 0$. Then

$$X^{-1}(-\infty, x] = \begin{cases} \emptyset & \text{if } x < 0, \\ \{\text{T}\} & \text{if } 0 \le x < 1, \\ \{\text{H, T}\} & \text{if } 1 \le x, \end{cases}$$

and we see that X is an RV.

Example 3. Let $\Omega = \{\text{HH, TT, HT, TH}\}$ and $\mathcal{S}$ be the class of all subsets of Ω. Define X by

$$X(\omega) = \text{number of H's in } \omega.$$

Then $X(\text{HH}) = 2$, $X(\text{HT}) = X(\text{TH}) = 1$, and $X(\text{TT}) = 0$.

$$X^{-1}(-\infty, x] = \begin{cases} \emptyset, & x < 0, \\ \{\text{TT}\}, & 0 \le x < 1, \\ \{\text{TT, HT, TH}\}, & 1 \le x < 2, \\ \Omega, & 2 \le x. \end{cases}$$

Thus X is an RV.

Remark 4. Let $(\Omega, \mathcal{S})$ be a discrete sample space; that is, let Ω be a countable set of points and $\mathcal{S}$ be the class of all subsets of Ω. Then every numerical-valued function defined on $(\Omega, \mathcal{S})$ is an RV.

Example 4. Let $\Omega = [0, 1]$ and $\mathcal{S} = \mathfrak{B} \cap [0, 1]$ be the σ-field of Borel sets on $[0, 1]$. Define X on Ω by

$$X(\omega) = \omega, \qquad \omega \in [0, 1].$$

Clearly, X is an RV. Any Borel subset of Ω is an event.

Remark 5. Let X be an RV defined on $(\Omega, \mathcal{S})$ and a, b be constants. Then $aX+b$ is also an RV on $(\Omega, \mathcal{S})$. Moreover, X^2 is an RV and so also is $1/X$, provided that $\{X = 0\} = \emptyset$. For a general result, see Theorem 2.5.1.

PROBLEMS 2.2

1. Let X be the number of heads in three tosses of a coin. What is Ω? What are the values that X assigns to points of Ω? What are the events $\{X \le 2.75\}$, $\{0.5 \le X \le 1.72\}$?

2. A die is tossed two times. Let X be the sum of face values on the two tosses and Y be the absolute value of the difference in face values. What is Ω? What values do X and Y assign to points of Ω? Check to see whether X and Y are random variables.

3. Let X be an RV. Is $|X|$ also an RV? If X is an RV that takes only nonnegative values, is $\sqrt{X}$ also an RV?

4. A die is rolled five times. Let X be the sum of face values. Write the events $\{X = 4\}$, $\{X = 6\}$, $\{X = 30\}$, and $\{X \geq 29\}$.

5. Let $\Omega = [0, 1]$ and $\mathcal{S}$ be the Borel σ-field of subsets of Ω. Define X on Ω as follows: $X(\omega) = \omega$ if $0 \leq \omega \leq \frac{1}{2}$ and $X(\omega) = \omega - \frac{1}{2}$ if $\frac{1}{2} < \omega \leq 1$. Is X an RV? If so, what is the event $\left\{\omega\colon X(\omega) \in \left(\frac{1}{4}, \frac{1}{2}\right)\right\}$?

6. Let $\mathfrak{A}$ be a class of subsets of $\mathcal{R}$ that generates $\mathfrak{B}$. Show that X is an RV on Ω if and only if $X^{-1}(A) \in \mathcal{R}$ for all $A \in \mathcal{A}$.

2.3 PROBABILITY DISTRIBUTION OF A RANDOM VARIABLE

In Section 2.2 we introduced the concept of an RV and noted that the concept of probability on the sample space was not used in this definition. In practice, however, random variables are of interest only when they are defined on a probability space. Let $(\Omega, \mathcal{S}, P)$ be a probability space, and let X be an RV defined on it.

Theorem 1. The RV X defined on the probability space $(\Omega, \mathcal{S}, P)$ induces a probability space $(\mathcal{R}, \mathfrak{B}, Q)$ by means of the correspondence

$$(1) \qquad Q(B) = P\{X^{-1}(B)\} = P\{\omega\colon X(\omega) \in B\} \qquad \text{for all } B \in \mathfrak{B}.$$

We write $Q = PX^{-1}$ and call Q or PX^{-1} the (probability) *distribution* of X.

Proof. Clearly, $Q(B) \geq 0$ for all $B \in \mathfrak{B}$, and also $Q(\mathcal{R}) = P\{X \in \mathcal{R}\} = P(\Omega) = 1$. Let $B_i \in \mathcal{B}$, $i = 1, 2, \ldots$, with $B_i \cap B_j = \emptyset$, $i \neq j$. Since the inverse image of a disjoint union of Borel sets is the disjoint union of their inverse images, we have

$$\begin{aligned} Q\left(\sum_{i=1}^{\infty} B_i\right) &= P\left\{X^{-1}\left(\sum_{i=1}^{\infty} B_i\right)\right\} \\ &= P\left\{\sum_{i=1}^{\infty} X^{-1}(B_i)\right\} \\ &= \sum_{i=1}^{\infty} PX^{-1}(B_i) = \sum_{i=1}^{\infty} Q(B_i). \end{aligned}$$

It follows that $(\mathcal{R}, \mathfrak{B}, Q)$ is a probability space, and the proof is complete.

We note that Q is a set function and that set functions are not easy to handle. It is therefore more practical to use (2.2.2) since then $Q(-\infty, x]$ is a point function. Let us first introduce and study some properties of a special point function on $\mathcal{R}$.

Definition 1. A real-valued function F defined on $(-\infty, \infty)$ that is nondecreasing, right continuous, and satisfies

$$F(-\infty) = 0 \quad \text{and} \quad F(+\infty) = 1$$

is called a *distribution function* (DF).

Remark 1. Recall that if F is a nondecreasing function on $\mathcal{R}$, then $F(x-) = \lim_{t\uparrow x} F(t)$, $F(x+) = \lim_{t\downarrow x} F(t)$ exist and are finite. Also, $F(+\infty)$ and $F(-\infty)$ exist as $\lim_{t\uparrow+\infty} F(t)$ and $\lim_{t\downarrow-\infty} F(t)$, respectively. In general,

$$F(x-) \le F(x) \le F(x+),$$

and x is a jump point of F if and only if $F(x+)$ and $F(x-)$ exist but are unequal. Thus a nondecreasing function F has only jump discontinuities. If we define

$$F^*(x) = F(x+) \qquad \text{for all } x,$$

we see that F^* is nondecreasing and right continuous on $\mathcal{R}$. Thus in Definition 1 the nondecreasing part is very important. Some authors demand left instead of right continuity in the definition of a DF.

Theorem 2. The set of discontinuity points of a DF F is at most countable.

Proof. Let $(a, b]$ be a finite interval with at least n discontinuity points:

$$a < x_1 < x_2 < \cdots < x_n \le b.$$

Then

$$F(a) \le F(x_1-) < F(x_1) \le \cdots \le F(x_n-) < F(x_n) \le F(b).$$

Let $p_k = F(x_k) - F(x_k-)$, $k = 1, 2, \ldots, n$. Clearly,

$$\sum_{k=1}^{n} p_k \le F(b) - F(a),$$

and it follows that the number of points x in $(a, b]$ with jump $p(x) > \varepsilon > 0$ is at most $\varepsilon^{-1}\{F(b) - F(a)\}$. Thus for every integer N, the number of discontinuity points with jump greater than $1/N$ is finite. It follows that there are no more than a countable number of discontinuity points in every finite interval $(a, b]$. Since $\mathcal{R}$ is a countable union of such intervals, the proof is complete.

Definition 2. Let X be an RV defined on $(\Omega, \mathcal{S}, P)$. Define a point function $F(\cdot)$ on $\mathcal{R}$ by using (1), namely,

$$F(x) = Q(-\infty, x] = P\{\omega\colon X(\omega) \le x\} \qquad \text{for all } x \in \mathcal{R}. \tag{2}$$

The function F is called the *distribution function* of RV X.

If there is no confusion, we will write

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

The following result justifies our calling F as defined by (2) a DF.

Theorem 3. The function F defined in (2) is indeed a DF.

Proof. Let $x_1 < x_2$. Then $(-\infty, x_1] \subset (-\infty, x_2]$, and we have

$$F(x_1) = P\{X \le x_1\} \le P\{X \le x_2\} = F(x_2).$$

Since F is nondecreasing, it is sufficient to show that for any sequence of numbers $x_n \downarrow x$, $x_1 > x_2 > \cdots > x_n > \cdots > x$, $F(x_n) \to F(x)$. Let $A_k = \{\omega\colon X(\omega) \in (x, x_k]\}$. Then $A_k \in \mathcal{S}$ and $A_k \not\uparrow$. Also,

$$\lim_{k\to\infty} A_k = \bigcap_{k=1}^{\infty} A_k = \emptyset,$$

since none of the intervals $(x, x_k]$ contains x. It follows that $\lim_{k\to\infty} P(A_k) = 0$. But

$$\begin{aligned} P(A_k) &= P\{X \le x_k\} - P\{X \le x\} \\ &= F(x_k) - F(x), \end{aligned}$$

so that

$$\lim_{k\to\infty} F(x_k) = F(x),$$

and F is right continuous.

Finally, let $\{x_n\}$ be a sequence of numbers decreasing to $-\infty$. Then

$$\{X \le x_n\} \supseteq \{X \le x_{n+1}\} \qquad \text{for each } n$$

and

$$\lim_{n\to\infty} \{X \le x_n\} = \bigcap_{n=1}^{\infty} \{X \le x_n\} = \emptyset.$$

Therefore,

$$F(-\infty) = \lim_{n\to\infty} P\{X \le x_n\} = P\left\{\lim_{n\to\infty}\{X \le x_n\}\right\} = 0.$$

Similarly,

$$F(+\infty) = \lim_{x_n\to\infty} P\{X \le x_n\} = 1,$$

and the proof is complete.

The next result, stated without proof, establishes a correspondence between the induced probability Q on $(\mathcal{R}, \mathfrak{B})$ and a point function F defined on $\mathcal{R}$.

Theorem 4. Given a probability Q on $(\mathcal{R}, \mathfrak{B})$, there exists a distribution function F satisfying

$$Q(-\infty, x] = F(x) \qquad \text{for all } x \in \mathcal{R}, \tag{3}$$

and conversely, given a DF F, there exists a unique probability Q defined on $(\mathcal{R}, \mathfrak{B})$ that satisfies (3).

For proof, see Chung [14, pp. 23–24].

Theorem 5. Every DF is the DF of an RV on some probability space.

Proof. Let F be a DF. From Theorem 4 it follows that there exists a unique probability Q defined on $\mathcal{R}$ that satisfies

$$Q(-\infty, x] = F(x) \qquad \text{for all } x \in \mathcal{R}.$$

Let $(\mathcal{R}, \mathfrak{B}, Q)$ be the probability space on which we define

$$X(\omega) = \omega, \qquad \omega \in \mathcal{R}.$$

Then

$$Q\{\omega\colon X(\omega) \le x\} = Q(-\infty, x] = F(x),$$

and F is the DF of RV X.

Remark 2. If X is an RV on $(\Omega, \mathcal{S}, P)$, we have seen (Theorem 3) that $F(x) = P\{X \le x\}$ is a DF associated with X. Theorem 5 assures us that to every DF F we can associate some RV. Thus, given an RV, there exists a DF, and conversely. In this book when we speak of an RV we will assume that it is defined on a probability space.

Example 1. Let X be defined on $(\Omega, \mathcal{S}, P)$ by

$$X(\omega) = c \qquad \text{for all } \omega \in \Omega.$$

Then

$$P\{X = c\} = 1,$$
$$F(x) = Q(-\infty, x] = P\{X^{-1}(-\infty, x]\} = 0 \qquad \text{if } x < c$$

and

$$F(x) = 1 \qquad \text{if } x \geq c.$$

Example 2. Let $\Omega = \{\text{H}, \text{T}\}$ and X be defined by

$$X(\text{H}) = 1, \qquad X(\text{T}) = 0.$$

If P assigns equal mass to $\{\text{H}\}$ and $\{\text{T}\}$, then

$$P\{X = 0\} = \tfrac{1}{2} = P\{X = 1\}$$

and

$$F(x) = Q(-\infty, x] = \begin{cases} 0, & x < 0, \\ \frac{1}{2}, & 0 \leq x < 1, \\ 1, & 1 \leq x. \end{cases}$$

Example 3. Let $\Omega = \{(i, j) \colon i, j \in \{1, 2, 3, 4, 5, 6\}\}$ and $\mathcal{S}$ be the set of all subsets of Ω. Let $P\{(i, j)\} = 1/6^2$ for all 6^2 pairs (i, j) in Ω. Define

$$X(i, j) = i + j, \qquad 1 \leq i, j \leq 6.$$

Then

$$F(x) = Q(-\infty, x] = P\{X \leq x\} = \begin{cases} 0, & x < 2, \\ \frac{1}{36}, & 2 \leq x < 3, \\ \frac{3}{36}, & 3 \leq x < 4, \\ \frac{6}{36}, & 4 \leq x < 5, \\ \vdots & \\ \frac{35}{36}, & 11 \leq x < 12, \\ 1, & 12 \leq x. \end{cases}$$

Example 4. We return to Example 2.2.4. For every subinterval I of $[0, 1]$, let $P(\text{I})$ be the length of the interval. Then $(\Omega, \mathcal{S}, P)$ is a probability space, and the DF of RV $X(\omega) = \omega$, $\omega \in \Omega$, is given by $F(x) = 0$ if $x < 0$, $F(x) = P\{\omega \colon X(\omega) \leq x\} = P([0, x]) = x$ if $x \in [0, 1]$, and $F(x) = 1$ if $x \geq 1$.

PROBLEMS 2.3

1. Write the DF of RV X defined in Problem 2.2.1, assuming that the coin is fair.

2. What is the DF of RV Y defined in Problem 2.2.2, assuming that the die is not loaded?

3. Do the following functions define DFs?
   (a) $F(x) = 0$ if $x < 0$, $= x$ if $0 \le x < \frac{1}{2}$, and $= 1$ if $x \ge \frac{1}{2}$.
   (b) $F(x) = (1/\pi)\tan^{-1} x$, $-\infty < x < \infty$.
   (c) $F(x) = 0$ if $x \le 1$, and $= 1 - (1/x)$ if $1 < x$.
   (d) $F(x) = 1 - e^{-x}$ if $x \ge 0$, and $= 0$ if $x < 0$.

4. Let X be an RV with DF F.
   (a) If F is the DF defined in Problem 3(a), find $P\{X > \frac{1}{4}\}$, $P\{\frac{1}{3} < X \le \frac{3}{8}\}$.
   (b) If F is the DF defined in Problem 3(d), find $P\{-\infty < X < 2\}$.

2.4 DISCRETE AND CONTINUOUS RANDOM VARIABLES

Let X be an RV defined on some fixed but otherwise arbitrary probability space $(\Omega, \mathcal{S}, P)$, and let F be the DF of X. In this book we restrict ourselves mainly to two cases: the case in which the RV assumes at most a countable number of values and hence its DF is a step function, and that in which the DF F is (absolutely) continuous.

Definition 1. An RV X defined on $(\Omega, \mathcal{S}, P)$ is said to be of the *discrete type*, or simply *discrete*, if there exists a countable set $E \subseteq \mathcal{R}$ such that $P\{X \in E\} = 1$. The points of E that have positive mass are called *jump points* or *points of increase* of the DF of X, and their probabilities are called *jumps* of the DF.

Note that $E \in \mathfrak{B}$ since every one-point set is in $\mathfrak{B}$. Indeed, if $x \in \mathcal{R}$, then

$$\{x\} = \bigcap_{n=1}^{\infty}\left[\left(x - \frac{1}{n} < x \le x + \frac{1}{n}\right)\right]. \tag{1}$$

Thus $\{X \in E\}$ is an event. Let X take on the value x_i with probability p_i ($i = 1, 2, \dots$). We have

$$P\{\omega\colon X(\omega) = x_i\} = p_i, \qquad i = 1, 2, \dots, \qquad p_i \ge 0 \text{ for all } i.$$

Then $\sum_{i=1}^{\infty} p_i = 1$.

Definition 2. The collection of numbers $\{p_i\}$ satisfying $P\{X = x_i\} = p_i \ge 0$, for all i and $\sum_{i=1}^{\infty} p_i = 1$, is called the *probability mass function* (PMF) of RV X.

The DF F of X is given by

$$F(x) = P\{X \le x\} = \sum_{x_i \le x} p_i. \tag{2}$$

If I_A denotes the indicator function of the set A, we may write

$$X(\omega) = \sum_{i=1}^{\infty} x_i I_{[X=x_i]}(\omega). \tag{3}$$

Let us define a function $\varepsilon(x)$ as follows:

$$\varepsilon(x) = \begin{cases} 1, & x \ge 0, \\ 0, & x < 0. \end{cases}$$

Then we have

$$F(x) = \sum_{i=1}^{\infty} p_i \varepsilon(x - x_i). \tag{4}$$

Example 1. The simplest example is that of an RV X *degenerate* at c, $P\{X = c\} = 1$:

$$F(x) = \varepsilon(x - c) = \begin{cases} 0, & x < c, \\ 1, & x \ge c. \end{cases}$$

Example 2. A box contains good and defective items. If an item drawn is good, we assign the number 1 to the drawing; otherwise, the number 0. Let p be the probability of drawing at random a good item. Then

$$P\left\{X = \begin{matrix} 0 \\ 1 \end{matrix}\right\} = \begin{cases} 1 - p, \\ p, \end{cases}$$

and

$$F(x) = P\{X \le x\} = \begin{cases} 0, & x < 0, \\ 1 - p, & 0 \le x < 1, \\ 1, & 1 \le x. \end{cases}$$

Example 3. Let X be an RV with PMF

$$P\{X = k\} = \frac{6}{\pi^2} \cdot \frac{1}{k^2}, \qquad k = 1, 2, \ldots.$$

Then

$$F(x) = \frac{6}{\pi^2} \sum_{k=1}^{\infty} \frac{1}{k^2} \varepsilon(x - k).$$

Theorem 1. Let $\{p_k\}$ be a collection of nonnegative real numbers such that $\sum_{k=1}^{\infty} p_k = 1$. Then $\{p_k\}$ is the PMF of some RV X.

We next consider RVs associated with DFs that have no jump points. The DF of such an RV is continuous. We shall restrict our attention to a special subclass of such RVs.

Definition 3. Let X be an RV defined on $(\Omega, \mathcal{S}, P)$ with DF F. Then X is said to be of the *continuous type* (or simply, *continuous*) if F is absolutely continuous, that is, if there exists a nonnegative function $f(x)$ such that for every real number x we have

$$F(x) = \int_{-\infty}^{x} f(t)\,dt. \tag{5}$$

The function f is called the *probability density function* (PDF) of the RV X.

Note that $f \geq 0$ and satisfies $\lim_{x\to+\infty} F(x) = F(+\infty) = \int_{-\infty}^{\infty} f(t)\,dt = 1$. Let a and b be any two real numbers with $a < b$. Then

$$\begin{aligned} P\{a < X \leq b\} &= F(b) - F(a) \\ &= \int_{a}^{b} f(t)\,dt. \end{aligned}$$

In view of remarks following Definition 2.2.1, the following result holds.

Theorem 2. Let X be an RV of the continuous type with PDF f. Then for every Borel set $B \in \mathfrak{B}$,

$$P(B) = \int_{B} f(t)\,dt. \tag{6}$$

If F is absolutely continuous and f is continuous at x, we have

$$F'(x) = \frac{dF(x)}{dx} = f(x). \tag{7}$$

Theorem 3. Every nonnegative real function f that is integrable over $\mathcal{R}$ and satisfies

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

is the PDF of some continuous RV X.

Proof. In view of Theorem 2.3.5, it suffices to show that there corresponds a DF F to f. Define

$$F(x) = \int_{-\infty}^{x} f(t)\,dt, \qquad x \in \mathcal{R}.$$

Then $F(-\infty) = 0$, $F(+\infty) = 1$, and if $x_2 > x_1$,

$$F(x_2) = \left(\int_{-\infty}^{x_1} + \int_{x_1}^{x_2}\right) f(t)\,dt \geq \int_{-\infty}^{x_1} f(t)\,dt = F(x_1).$$

Finally, F is (absolutely) continuous and hence continuous from the right.

Remark 1. In the discrete case, $P\{X = a\}$ is the probability that X takes the value a. In the continuous case, $f(a)$ is not the probability that X takes the value a. Indeed, if X is of the continuous type, it assumes every value with probability 0.

Theorem 4. Let X be any RV. Then

(8) $$P\{X = a\} = \lim_{\substack{t \to a \\ t < a}} P\{t < X \leq a\}.$$

Proof. Let $t_1 < t_2 < \cdots < a$, $\quad t_n \to a$, and write

$$A_n = \{t_n < X \leq a\}.$$

Then A_n is a nonincreasing sequence of events that converges to $\bigcap_{n=1}^{\infty} A_n = \{X = a\}$. It follows that $\lim_{n\to\infty} PA_n = P\{X = a\}$.

Remark 2. Since $P\{t < X \leq a\} = F(a) - F(t)$, it follows that

$$\begin{aligned}\lim_{\substack{t \to a \\ t < a}} P\{t < X \leq a\} = P\{X = a\} &= F(a) - \lim_{\substack{t \to a \\ t < a}} F(t)\\ &= F(a) - F(a-).\end{aligned}$$

Thus F has a jump discontinuity at a if and only if $P\{X = a\} > 0$; that is, F is continuous at a if and only if $P\{X = a\} = 0$. If X is an RV of the continuous type, $P\{X = a\} = 0$ for all $a \in \mathcal{R}$. Moreover,

$$P\{X \in \mathcal{R} - \{a\}\} = 1.$$

This justifies Remark 1.3.4.

Remark 3. The set of real numbers x for which a DF F increases is called the *support* of F. Let X be the RV with DF F, and let S be the support of F. Then

$P(X \in S) = 1$ and $P(X \in S^c) = 0$. The set of positive integers is the support of the DF in Example 3, and the open interval (0, 1) is the support of F in Example 4.

Example 4. Let X be an RV with DF F given by (Fig. 1)

$$F(x) = \begin{cases} 0, & x \leq 0, \\ x, & 0 < x \leq 1, \\ 1, & 1 < x. \end{cases}$$

Differentiating F with respect to x at continuity points of f, we get

$$f(x) = F'(x) = \begin{cases} 0, & x < 0 \text{ or } x > 1, \\ 1, & 0 < x < 1. \end{cases}$$

The function f is not continuous at $x = 0$ or at $x = 1$ (Fig. 2). We may define $f(0)$ and $f(1)$ in any manner. Choosing $f(0) = f(1) = 0$, we have

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

Then

$$P\{0.4 < X \leq 0.6\} = F(0.6) - F(0.4) = 0.2.$$

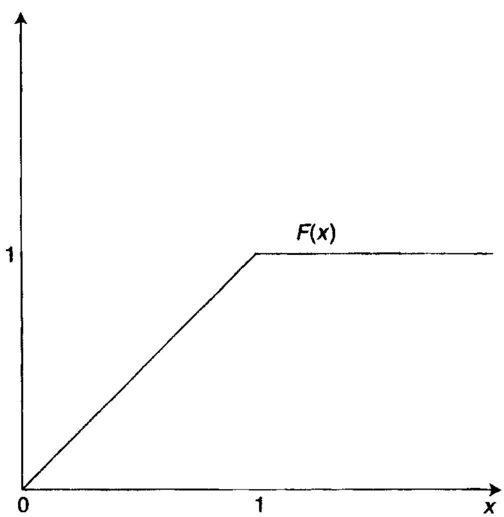

Fig. 1.

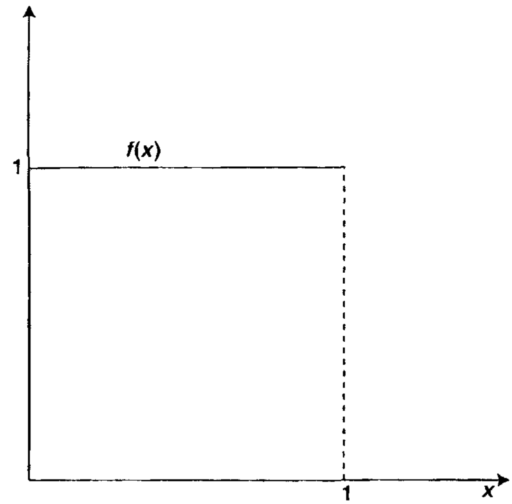

Fig. 2.

Example 5. Let X have the *triangular* PDF (Fig. 3)

$$f(x) = \begin{cases} x, & 0 < x \le 1, \\ 2 - x, & 1 \le x \le 2, \\ 0, & \text{otherwise}. \end{cases}$$

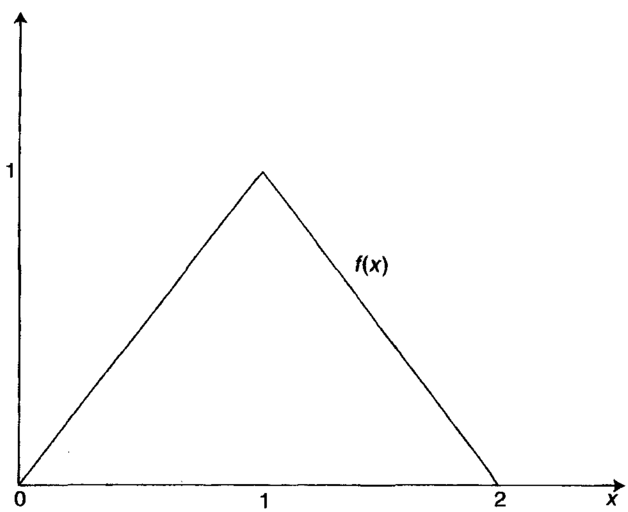

Fig. 3. Graph of f.

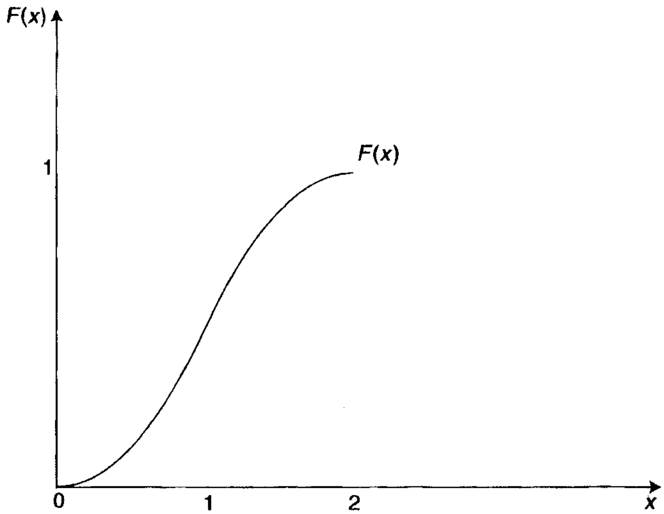

Fig. 4. Graph of F.

It is easy to check that f is a PDF. For the DF F of X we have (Fig. 4)

$$F(x) = \begin{cases} 0 & \text{if } x \le 0, \\ \displaystyle\int_0^x t\,dt = \frac{x^2}{2} & \text{if } 0 < x \le 1, \\ \displaystyle\int_0^1 t\,dt + \int_1^x (2-t)\,dt = 2x - \frac{x^2}{2} - 1 & \text{if } 1 < x \le 2, \\ 1 & \text{if } x \ge 2. \end{cases}$$

Then

$$\begin{aligned} P\{0.3 < X \le 1.5\} &= P\{X \le 1.5\} - P\{X \le 0.3\} \\ &= 0.83. \end{aligned}$$

Example 6. Let $k > 0$ be a constant, and

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

Then $\int_0^1 f(x)\,dx = k/6$. It follows that $f(x)$ defines a PDF if $k = 6$. We have

$$P\{X > 0.3\} = 1 - 6\int_0^3 x(1-x)\,dx = 0.784.$$

We conclude this discussion by emphasizing that the two types of RVs considered above form only a part of the class of all RVs. These two classes, however, contain practically all the random variables that arise in practice. We note without proof (see Chung [14, p. 9]) that every DF F can be decomposed into two parts according to

$$F(x) = aF_d(x) + (1-a)F_c(x). \tag{9}$$

Here F_d and F_c are both DFs; F_d is the DF of a discrete RV, while F_c is a continuous (not necessarily absolutely continuous) DF. In fact, F_c can be further decomposed, but we will not go into that (see Chung [14, p. 11]).

Example 7. Let X be an RV with DF

$$F(x) = \begin{cases} 0, & x < 0, \\ \dfrac{1}{2}, & x = 0, \\ \dfrac{1}{2} + \dfrac{x}{2}, & 0 < x < 1, \\ 1, & 1 \le x. \end{cases}$$

Note that the DF F has a jump at $x = 0$ and F is continuous (in fact, absolutely continuous) in the interval $(0, 1)$. F is the DF of an RV X that is neither discrete nor continuous. We can write

$$F(x) = \tfrac{1}{2}F_d(x) + \tfrac{1}{2}F_c(x),$$

where

$$F_d(x) = \begin{cases} 0, & x < 0, \\ 1, & x \ge 0; \end{cases}$$

and

$$F_c(x) = \begin{cases} 0, & x \le 0, \\ x, & 0 < x < 1, \\ 1, & 1 \le x. \end{cases}$$

Here $F_d(x)$ is the DF of the RV degenerate at $x = 0$, and $F_c(x)$ is the DF with PDF

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

PROBLEMS 2.4

1. Let

$$p_k = p(1-p)^k, \qquad k = 0, 1, 2, \ldots, \qquad 0 < p < 1.$$

Does $\{p_k\}$ define the PMF of some RV? What is the DF of this RV? If X is an RV with PMF $\{p_k\}$, what is $P\{n \le X \le N\}$, where n, N $(N > n)$ are positive integers?

2. In Problem 2.3.3, find the PDF associated with the DFs of parts (b), (c), and (d).

3. Does the function $f_\theta(x) = \theta^2 x e^{-\theta x}$ if $x > 0$, and $= 0$ if $x \le 0$, where $\theta > 0$, define a PDF? Find the DF associated with $f_\theta(x)$; if X is an RV with PDF $f_\theta(x)$, find $P\{X \ge 1\}$.

4. Does the function $f_\theta(x) = \{(x+1)/[\theta(\theta+1)]\}e^{-x/\theta}$ if $x > 0$, and $= 0$ otherwise, where $\theta > 0$ define a PDF? Find the corresponding DF.

5. For what values of K do the following functions define the PMF of some RV?
   (a) $f(x) = K(\lambda^x/x!)$, $x = 0, 1, 2, \ldots$, $\lambda > 0$.
   (b) $f(x) = K/N$, $x = 1, 2, \ldots, N$.

6. Show that the function

$$f(x) = \tfrac{1}{2}e^{-|x|}, \qquad -\infty < x < \infty,$$

is a PDF. Find its DF.

7. For the PDF $f(x) = x$ if $0 \le x < 1$, and $= 2 - x$ if $1 \le x < 2$, find $P\{\frac{1}{6} < X \le \frac{7}{4}\}$.

8. Which of the following functions are density functions?
   (a) $f(x) = x(2-x)$, $0 < x < 2$, and 0 elsewhere.
   (b) $f(x) = x(2x-1)$, $0 < x < 2$, and 0 elsewhere.
   (c) $f(x) = (1/\lambda)\exp\{-(x-\theta)/\lambda\}$, $x > \theta$, and 0 elsewhere, $\lambda > 0$.
   (d) $f(x) = \sin x$, $0 < x < \pi/2$, and 0 elsewhere.
   (e) $f(x) = 0$ for $x < 0$, $= (x+1)/9$ for $0 \le x < 1$, $= 2(2x-1)/9$ for $1 \le x < \frac{3}{2}$, $= 2(5-2x)/9$ for $\frac{3}{2} \le x < 1$, $= \frac{4}{27}$ for $2 \le x < 5$, and 0 elsewhere.
   (f) $f(x) = 1/[\pi(1+x^2)]$, $x \in \mathcal{R}$.

9. Are the following functions distribution functions? If so, find the corresponding density or probability functions.
   (a) $F(x) = 0$ for $x \le 0$, $= x/2$ for $0 \le x < 1$, $= \frac{1}{2}$ for $1 \le x < 2$, $= x/4$ for $2 \le x < 4$ and $= 1$ for $x \ge 4$.
   (b) $F(x) = 0$ if $x < -\theta$, $= \frac{1}{2}(x/\theta + 1)$ if $|x| \le \theta$, and 1 for $x > \theta$ where $\theta > 0$.

(c) $F(x) = 0$ if $x < 0$, and $= 1 - (1+x)\exp(-x)$ if $x \geq 0$.

(d) $F(x) = 0$ if $x < 1$, $= (x-1)^2/8$ if $1 \leq x < 3$, and 1 for $x \geq 3$.

(e) $F(x) = 0$ if $x < 0$, and $= 1 - e^{-x^2}$ if $x \geq 0$.

10. Suppose that $P(X \geq x)$ is given for a random variable X (of the continuous type) for all x. How will you find the corresponding density function? In particular, find the density function in each of the following cases:

(a) $P(X \geq x) = 1$ if $x \leq 0$, and $P(X \geq x) = e^{-\lambda x}$ for $x > 0$; $\lambda > 0$ is a constant.

(b) $P(X \geq x) = 1$ if $x < 0$, and $= (1+x/\lambda)^{-\lambda}$, for $x \geq 0$, $\lambda > 0$ is a constant.

(c) $P(X \geq x) = 1$ if $x \leq 0$, and $= 3/(1+x)^2 - 2/(1+x)^3$ if $x > 0$.

(d) $P(X > x) = 1$ if $x \leq x_0$, and $= (x_0/x)^\alpha$ if $x > x_0$; $x_0 > 0$ and $\alpha > 0$ are constants.

2.5 FUNCTIONS OF A RANDOM VARIABLE

Let X be an RV with a known distribution, and let g be a function defined on the real line. We seek the distribution of $Y = g(X)$, provided that Y is also an RV. We first prove the following result.

Theorem 1. Let X be an RV defined on $(\Omega, \mathcal{S}, P)$. Also, let g be a Borel-measurable function on $\mathcal{R}$. Then $g(X)$ is also an RV.

Proof. For $y \in \mathcal{R}$, we have

$$\{g(X) \leq y\} = \{X \in g^{-1}(-\infty, y]\},$$

and since g is Borel-measurable, $g^{-1}(-\infty, y]$ is a Borel set. It follows that $\{g(X) \leq y\} \in \mathcal{S}$, and the proof is complete.

Theorem 2. Given an RV X with a known DF, the distribution of the RV $Y = g(X)$, where g is a Borel-measurable function, is determined.

Proof. Indeed, for all $y \in \mathcal{R}$,

$$P\{Y \leq y\} = P\{X \in g^{-1}(-\infty, y]\}. \tag{1}$$

In what follows we always assume that the functions under consideration are Borel-measurable.

Example 1. Let X be an RV with DF F. Then $|X|$, $aX + b$ (where $a \neq 0$ and b are constants), X^k (where $k \geq 0$ is an integer), and $|X|^\alpha$ ($\alpha > 0$) are all RVs. Define

$$X^+ = \begin{cases} X, & X \geq 0, \\ 0, & X < 0, \end{cases}$$

and

$$X^- = \begin{cases} X, & X \le 0, \\ 0, & X > 0. \end{cases}$$

Then X^+, X^- are also RVs. We have

$$\begin{aligned} P\{|X| \le y\} &= P\{-y \le X \le y\} = P\{X \le y\} - P\{X < -y\} \\ &= F(y) - F(-y) + P\{X = -y\}, \qquad y > 0; \\ P\{aX + b \le y\} &= P\{aX \le y - b\} \\ &= \begin{cases} P\left\{X \le \dfrac{y-b}{a}\right\} & \text{if } a > 0, \\ P\left\{X \ge \dfrac{y-b}{a}\right\} & \text{if } a < 0; \end{cases} \end{aligned}$$

and

$$P\{X^+ \le y\} = \begin{cases} 0 & \text{if } y < 0, \\ P\{X \le 0\} & \text{if } y = 0, \\ P\{X < 0\} + P\{0 \le X \le y\} & \text{if } y > 0. \end{cases}$$

Similarly,

$$P\{X^- \le y\} = \begin{cases} 1 & \text{if } y \ge 0, \\ P\{X \le y\} & \text{if } y < 0. \end{cases}$$

Let X be an RV of the discrete type and A be the countable set such that $P\{X \in A\} = 1$ and $P\{X = x\} > 0$ for $x \in A$. Let $Y = g(X)$ be a one-to-one mapping from A onto some set B. Then the inverse map, g^{-1}, is a single-valued function of y. To find $P\{Y = y\}$, we note that

$$P\{Y = y\} = \begin{cases} P\{g(X) = y\} = P\{X = g^{-1}(y)\}, & y \in B, \\ 0, & y \in B^c. \end{cases}$$

Example 2. Let X be a *Poisson* RV with PMF

$$P\{X = k\} = \begin{cases} e^{-\lambda}\dfrac{\lambda^k}{k!}, & k = 0, 1, 2, \ldots; \quad \lambda > 0, \\ 0, & \text{otherwise}. \end{cases}$$

Let $Y = X^2 + 3$. Then $y = x^2 + 3$ maps $A = \{0, 1, 2, \ldots\}$ onto $B = \{3, 4, 7, 12, 19, 28, \ldots\}$. The inverse map is $x = \sqrt{y - 3}$, and since there are no negative values in A, we take the positive square root of $y-3$. We have

$$P\{Y=y\}=P\{X=\sqrt{y-3}\}=\frac{e^{-\lambda}\lambda^{\sqrt{y-3}}}{\sqrt{(y-3)}!}, \qquad y\in B,$$

and $P\{Y=y\}=0$ elsewhere.

Actually, the restriction to a single-valued inverse on g is not necessary. If g has a finite (or even a countable) number of inverses for each y, from countable additivity of P we have

$$\begin{aligned}P\{Y=y\}=P\{g(X)=y\}&=P\left\{\bigcup_a [X=a, g(a)=y]\right\}\\&=\sum_a P\{X=a, g(a)=y\}.\end{aligned}$$

Example 3. Let X be an RV with PMF

$$P\{X=-2\}=\tfrac{1}{5},\quad P\{X=-1\}=\tfrac{1}{6},\quad P\{X=0\}=\tfrac{1}{5},$$
$$P\{X=1\}=\tfrac{1}{15},\quad\text{and}\quad P\{X=2\}=\tfrac{11}{30}.$$

Let $Y=X^2$. Then

$$A=\{-2,-1,0,1,2\}\quad\text{and}\quad B=\{0,1,4\}.$$

We have

$$P\{Y=y\}=\begin{cases}\frac{1}{5} & y=0,\\ \frac{1}{6}+\frac{1}{15}=\frac{7}{30}, & y=1,\\ \frac{1}{5}+\frac{11}{30}=\frac{17}{30}, & y=4.\end{cases}$$

The case in which X is an RV of the continuous type is not as simple. First we note that, if X is a continuous RV and g is some Borel-measurable function, $Y=g(X)$ may not be an RV of the continuous type.

Example 4. Let X be an RV with *uniform* distribution on $[-1,1]$; that is, the PDF of X is $f(x)=\frac{1}{2}$, $-1\le x\le 1$, and $=0$ elsewhere. Let $Y=X^+$. Then, from Example 1,

$$P\{Y\le y\}=\begin{cases}0, & y<0,\\ \frac{1}{2}, & y=0,\\ \frac{1}{2}+\frac{1}{2}y, & 1\ge y>0,\\ 1, & y>1.\end{cases}$$

We see that the DF of Y has a jump at $y = 0$ and that Y is neither discrete nor continuous. Note that all we require is that $P\{X < 0\} > 0$ for X^+ to be of the mixed type.

Example 4 shows that we need some conditions on g to ensure that $g(X)$ is also an RV of the continuous type whenever X is continuous. This is the case when g is a continuous monotonic function. A sufficient condition is given in the following theorem.

Theorem 3. Let X be an RV of the continuous type with PDF f. Let $y = g(x)$ be differentiable for all x and either $g'(x) > 0$ for all x or $g'(x) < 0$ for all x. Then $Y = g(X)$ is also an RV of the continuous type with PDF given by

$$\text{(2)} \qquad h(y) = \begin{cases} f[g^{-1}(y)] \left| \dfrac{d}{dy} g^{-1}(y) \right|, & \alpha < y < \beta, \\ 0, & \text{otherwise,} \end{cases}$$

where $\alpha = \min\{g(-\infty), g(+\infty)\}$ and $\beta = \max\{g(-\infty), g(+\infty)\}$.

Proof. If g is differentiable for all x and $g'(x) > 0$ for all x, then g is continuous and strictly increasing, the limits α, β exist (may be infinite), and the inverse function $x = g^{-1}(y)$ exists, is strictly increasing, and is differentiable. The DF of Y for $\alpha < y < \beta$ is given by

$$P\{Y \le y\} = P\{X \le g^{-1}(y)\}.$$

The PDF of g is obtained on differentiation. We have

$$\begin{aligned} h(y) &= \frac{d}{dy} P\{Y \le y\} \\ &= f[g^{-1}(y)] \frac{d}{dy} g^{-1}(y). \end{aligned}$$

Similarly, if $g' < 0$, then g is strictly decreasing and we have

$$\begin{aligned} P\{Y \le y\} &= P\{X \ge g^{-1}(y)\} \\ &= 1 - P\{X \le g^{-1}(y)\} \qquad (X \text{ is a continuous RV}) \end{aligned}$$

so that

$$h(y) = -f[g^{-1}(y)] \cdot \frac{d}{dy} g^{-1}(y).$$

Since g and g^{-1} are both strictly decreasing, $(d/dy)\, g^{-1}(y)$ is negative and (2) follows.

Note that

$$\frac{d}{dy}g^{-1}(y) = \left.\frac{1}{dg(x)/dx}\right|_{x=g^{-1}(y)},$$

so that (2) may be rewritten as

$$h(y) = \left.\frac{f(x)}{|dg(x)/dx|}\right|_{x=g^{-1}(y)}, \qquad \alpha < y < \beta. \tag{3}$$

Remark 1. The key to computation of the induced distribution of $Y = g(X)$ from the distribution of X is (1). If the conditions of Theorem 3 are satisfied, we are able to identify the set $\{X \in g^{-1}(-\infty, y]\}$ as $\{X \le g^{-1}(y)\}$ or $\{X \ge g^{-1}(y)\}$, according to whether g is increasing or decreasing. In practice, Theorem 3 is quite useful, but whenever the conditions are violated, one should return to (1) to compute the induced distribution. This is the case, for example, in Examples 7 and 8 and Theorem 4 below.

Remark 2. If the PDF f of X vanishes outside an interval $[a, b]$ of finite length, we need only to assume that g is differentiable in (a, b), and either $g'(x) > 0$ or $g'(x) < 0$ throughout the interval. Then we take

$$\alpha = \min\{g(a), g(b)\} \quad \text{and} \quad \beta = \max\{g(a), g(b)\}$$

in Theorem 3.

Example 5. Let X have the density $f(x) = 1$, $0 < x < 1$, and $= 0$ otherwise. Let $Y = e^X$. Then $X = \log Y$, and we have

$$h(y) = \left|\frac{1}{y}\right| \cdot 1, \qquad 0 < \log y < 1,$$

that is,

$$h(y) = \begin{cases} \dfrac{1}{y}, & 1 < y < e, \\ 0, & \text{otherwise.} \end{cases}$$

If $y = -2\log x$, then $x = e^{-y/2}$ and

$$\begin{aligned} h(y) &= |-\tfrac{1}{2}e^{-y/2}| \cdot 1, \qquad & 0 < e^{-y/2} < 1, \\ &= \begin{cases} \tfrac{1}{2}e^{-y/2}, \\ 0, \end{cases} & \begin{array}{l} 0 < y < \infty, \\ \text{otherwise.} \end{array} \end{aligned}$$

Example 6. Let X be a nonnegative RV of the continuous type with PDF f, and let $\alpha > 0$. Let $Y = X^\alpha$. Then

$$P\{X^\alpha \le y\} = \begin{cases} P\{X \le y^{1/\alpha}\} & \text{if } y \ge 0, \\ 0 & \text{if } y < 0. \end{cases}$$

The PDF of Y is given by

$$\begin{aligned} h(y) &= f(y^{1/\alpha}) \left| \frac{d}{dy} y^{1/\alpha} \right| \\ &= \begin{cases} \dfrac{1}{\alpha} y^{1/\alpha - 1} f(y^{1/\alpha}), & y > 0, \\ 0, & y \le 0. \end{cases} \end{aligned}$$

Example 7. Let X be an RV with PDF

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \qquad -\infty < x < \infty.$$

Let $Y = X^2$. In this case, $g'(x) = 2x$, which is > 0 for $x > 0$, and < 0 for $x < 0$, so that the conditions of Theorem 3 are not satisfied. But for $y > 0$,

$$\begin{aligned} P\{Y \le y\} &= P\{-\sqrt{y} \le X \le \sqrt{y}\} \\ &= F(\sqrt{y}) - F(-\sqrt{y}), \end{aligned}$$

where F is the DF of X. Thus the PDF of Y is given by

$$h(y) = \begin{cases} \dfrac{1}{2\sqrt{y}} [f(\sqrt{y}) + f(-\sqrt{y})], & y > 0, \\ 0, & y \le 0. \end{cases}$$

Thus

$$h(y) = \begin{cases} \dfrac{1}{\sqrt{2\pi\, y}} e^{-y/2}, & 0 < y, \\ 0, & y \le 0. \end{cases}$$

Example 8. Let X be an RV with PDF

$$f(x) = \begin{cases} \dfrac{2x}{\pi^2}, & 0 < x < \pi, \\ 0, & \text{otherwise}. \end{cases}$$

Let $Y = \sin X$. In this case $g'(x) = \cos x > 0$ for x in $(0, \pi/2)$ and < 0 for x in $(\pi/2, \pi)$, so that the conditions of Theorem 3 are not satisfied. To compute the PDF

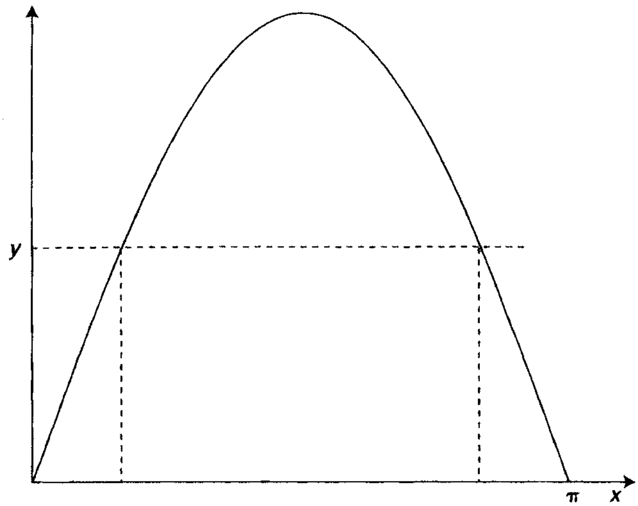

Fig. 1. $y = \sin x,\ 0 \le x \le \pi.$

of Y, we return to (1) and see that (Fig. 1) the DF of Y is given by

$$\begin{aligned} P\{Y \le y\} &= P\{\sin X \le y\}, \qquad 0 < y < 1, \\ &= P\{(0 \le X \le x_1) \cup (x_2 \le X \le \pi)\}, \end{aligned}$$

where $x_1 = \sin^{-1} y$ and $x_2 = \pi - \sin^{-1} y$. Thus

$$\begin{aligned} P\{Y \le y\} &= \int_0^{x_1} f(x)\,dx + \int_{x_2}^{\pi} f(x)\,dx \\ &= \left(\frac{x_1}{\pi}\right)^2 + 1 - \left(\frac{x_2}{\pi}\right)^2, \end{aligned}$$

and the PDF of Y is given by

$$\begin{aligned} h(y) &= \frac{d}{dy}\left(\frac{\sin^{-1} y}{\pi}\right)^2 + \frac{d}{dy}\left[1 - \left(\frac{\pi - \sin^{-1} y}{\pi}\right)^2\right] \\ &= \begin{cases} \dfrac{2}{\pi\sqrt{1-y^2}}, & 0 < y < 1, \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

In Examples 7 and 8 the function $y = g(x)$ can be written as the sum of two monotone functions. We applied Theorem 3 to each of these monotonic summands. These two examples are special cases of the following result.

Theorem 4. Let X be an RV of the continuous type with PDF f. Let $y = g(x)$ be differentiable for all x, and assume that $g'(x)$ is continuous and nonzero at all but a finite number of values of x. Then for every real number y,

(a) there exist a positive integer $n = n(y)$ and real numbers (inverses) $x_1(y)$, $x_2(y), \dots, x_n(y)$ such that

$$g[x_k(y)] = y \quad \text{and} \quad g'[x_k(y)] \neq 0, \qquad k = 1, 2, \dots, n(y),$$

or

(b) there does not exist any x such that $g(x) = y$, $g'(x) \neq 0$, in which case we write $n(y) = 0$.

Then Y is a continuous RV with PDF given by

$$h(y) = \begin{cases} \displaystyle\sum_{k=1}^{n} f[x_k(y)]\,|g'[x_k(y)]|^{-1} & \text{if } n > 0, \\ 0 & \text{if } n = 0. \end{cases}$$

Example 9. Let X be an RV with PDF f, and let $Y = |X|$. Here $n(y) = 2$, $x_1(y) = y$, $x_2(y) = -y$ for $y > 0$, and

$$h(y) = \begin{cases} f(y) + f(-y), & y > 0, \\ 0, & y \le 0. \end{cases}$$

Thus, if $f(x) = \frac{1}{2}$, $-1 \le x \le 1$, and $= 0$ otherwise, then

$$h(y) = \begin{cases} 1, & 0 \le y \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

If $f(x) = (1/\sqrt{2\pi})e^{-(x^2/2)}$, $-\infty < x < \infty$, then

$$h(y) = \begin{cases} \dfrac{2}{\sqrt{2\pi}} e^{-(y^2/2)}, & y > 0, \\ 0, & \text{otherwise.} \end{cases}$$

Example 10. Let X be an RV of the continuous type with PDF f, and let $Y = X^{2m}$, where m is a positive integer. In this case $g(x) = x^{2m}$, $g'(x) = 2mx^{2m-1} > 0$ for $x > 0$ and $g'(x) < 0$ for $x < 0$. Writing $n = 2m$, we see that for any $y > 0$, $n(y) = 2$, $x_1(y) = -y^{1/n}$, $x_2(y) = y^{1/n}$. It follows that

$$\begin{aligned} h(y) &= f[x_1(y)] \cdot \frac{1}{ny^{1-1/n}} + f[x_2(y)]\frac{1}{ny^{1-1/n}} \\ &= \begin{cases} \dfrac{1}{ny^{1-1/n}}[f(y^{1/n}) + f(-y^{1/n})] & \text{if } y > 0, \\ 0 & \text{if } y \le 0. \end{cases} \end{aligned}$$

In particular, if f is the PDF given in Example 7, then

$$h(y) = \begin{cases} \dfrac{2}{\sqrt{2\pi}\, n y^{1-1/n}} \exp\left(-\dfrac{y^{2/n}}{2}\right) & \text{if } y > 0, \\ 0 & \text{if } y \le 0. \end{cases}$$

Remark 3. The basic formula (1) and the countable additivity of probability allow us to compute the distribution of $Y = g(X)$ in some instances even if g has a countable number of inverses. Let $A \subseteq \mathcal{R}$ and g map A into $B \subseteq \mathcal{R}$. Suppose that A can be represented as a countable union of disjoint sets A_k, $k = 1, 2, \ldots$. Then the DF of Y is given by

$$\begin{aligned} P\{Y \le y\} &= P\{X \in g^{-1}(-\infty, y]\} \\ &= P\left\{X \in \sum_{k=1}^{\infty} [\{g^{-1}(-\infty, y]\} \cap A_k]\right\} \\ &= \sum_{k=1}^{\infty} P\left\{X \in A_k \cap \{g^{-1}(-\infty, y]\}\right\}. \end{aligned}$$

If the conditions of Theorem 3 are satisfied by the restriction of g to each A_k, we may obtain the PDF of Y on differentiating the DF of Y. We remind the reader that term-by-term differentiation is permissible if the differentiated series is uniformly convergent.

Example 11. Let X be an RV with PDF

$$f(x) = \begin{cases} \theta e^{-\theta x}, & x > 0, \\ 0, & x \le 0, \end{cases} \qquad \theta > 0.$$

Let $Y = \sin X$, and let $\sin^{-1} y$ be the principal value. Then (Fig. 2) for $0 < y < 1$,

$$\begin{aligned} &P\{\sin X \le y\} \\ &\quad = P\{0 < X \le \sin^{-1} y \text{ or } (2n-1)\pi - \sin^{-1} y \le X \le 2n\pi + \sin^{-1} y \\ &\qquad\qquad \text{for all integers } n \ge 1\} \\ &\quad = P\{0 < X \le \sin^{-1} y\} + \sum_{n=1}^{\infty} P\{(2n-1)\pi - \sin^{-1} y \le X \le 2n\pi + \sin^{-1} y\} \\ &\quad = 1 - e^{-\theta \sin^{-1} y} + \sum_{n=1}^{\infty} \left(e^{-\theta[(2n-1)\pi - \sin^{-1} y]} - e^{-\theta(2n\pi + \sin^{-1} y)}\right) \end{aligned}$$

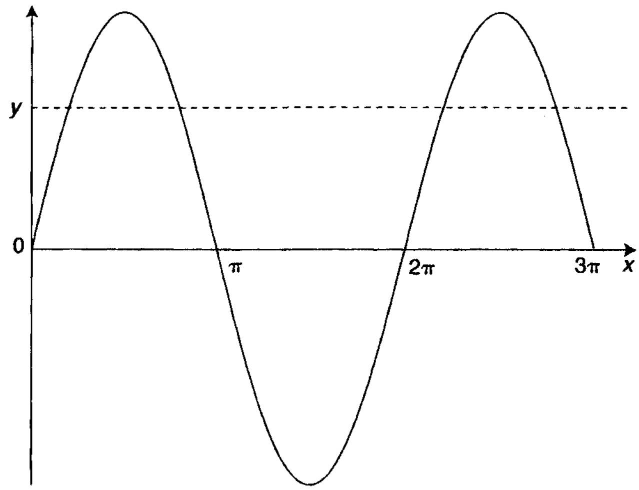

Fig. 2. $y = \sin x, \; x \geq 0.$

$$= 1 - e^{-\theta \sin^{-1} y} + (e^{\theta\pi + \theta \sin^{-1} y} - e^{-\theta \sin^{-1} y}) \sum_{n=1}^{\infty} e^{-(2\theta\pi)n}$$

$$= 1 - e^{-\theta \sin^{-1} y} + (e^{\theta\pi + \theta \sin^{-1} y} - e^{-\theta \sin^{-1} y}) \frac{e^{-2\theta\pi}}{1 - e^{-2\theta\pi}}$$

$$= 1 + \frac{e^{-\theta\pi + \theta \sin^{-1} y} - e^{-\theta \sin^{-1} y}}{1 - e^{-2\pi\theta}}.$$

A similar computation can be made for $y < 0$. It follows that the PDF of Y is given by

$$h(y) = \begin{cases} \theta e^{-\theta\pi}(1 - e^{-2\theta\pi})^{-1}(1 - y^2)^{-1/2}\left(e^{\theta \sin^{-1} y} + e^{-\theta\pi - \theta \sin^{-1} y}\right) & \text{if } -1 < y < 0, \\ \theta(1 - e^{-2\theta\pi})^{-1}(1 - y^2)^{-1/2}\left(e^{-\theta \sin^{-1} y} + e^{-\theta\pi + \theta \sin^{-1} y}\right) & \text{if } 0 < y < 1, \\ 0 & \text{otherwise.} \end{cases}$$

PROBLEMS 2.5

1. Let X be a random variable with probability mass function

$$P\{X = r\} = \binom{n}{r} p^r (1 - p)^{n-r}, \qquad r = 0, 1, 2, \ldots, n, \quad 0 \leq p \leq 1.$$

Find the PMFs of the RVs (a) $Y = aX + b$, (b) $Y = X^2$, and (c) $Y = \sqrt{X}$.

2. Let X be an RV with PDF

$$f(x) = \begin{cases} 0 & \text{if } x \le 0, \\ \dfrac{1}{2} & \text{if } 0 < x \le 1, \\ \dfrac{1}{2x^2} & \text{if } 1 < x < \infty. \end{cases}$$

Find the PDF of the RV $1/X$.

3. Let X be a positive RV of the continuous type with PDF $f(\cdot)$. Find the PDF of the RV $U = X/(1+X)$. If, in particular, X has the PDF

$$f(x) = \begin{cases} 1, & 0 \le x \le 1, \\ 0, & \text{otherwise,} \end{cases}$$

what is the PDF of U?

4. Let X be an RV with PDF f defined by Example 11. Let $Y = \cos X$ and $Z = \tan X$. Find the DFs and PDFs of Y and Z.

5. Let X be an RV with PDF

$$f_\theta(x) = \begin{cases} \theta e^{-\theta x} & \text{if } x \ge 0, \\ 0 & \text{otherwise,} \end{cases}$$

where $\theta > 0$. Let $Y = (X - 1/\theta)^2$. Find the PDF of Y.

6. A point is chosen at random on the circumference of a circle of radius r with center at the origin, that is, the polar angle θ of the point chosen has the PDF

$$f(\theta) = \frac{1}{2\pi}, \qquad \theta \in (-\pi, \pi).$$

Find the PDF of the abscissa of the point selected.

7. For the RV X of Example 7, find the PDF of the following RVs: (a) $Y_1 = e^X$, (b) $Y_2 = 2X^2 + 1$, and (c) $Y_3 = g(X)$, where $g(x) = 1$ if $x > 0$, $= \frac{1}{2}$ if $x = 0$, and $= -1$ if $x < 0$.

8. Suppose that a projectile is fired at an angle θ above the earth with a velocity V. Assuming that θ is an RV with PDF

$$f(\theta) = \begin{cases} \dfrac{12}{\pi} & \text{if } \dfrac{\pi}{6} < \theta < \dfrac{\pi}{4}, \\ 0 & \text{otherwise,} \end{cases}$$

find the PDF of the range R of the projectile, where $R = V^2 \sin 2\theta / g$, g being the gravitational constant.

9. Let X be an RV with PDF $f(x) = 1/(2\pi)$ if $0 < x < 2\pi$, and $= 0$ otherwise. Let $Y = \sin X$. Find the DF and PDF of Y.

10. Let X be an RV with PDF $f(x) = \frac{1}{3}$ if $-1 < x < 2$, and $= 0$ otherwise. Let $Y = |X|$. Find the PDF of Y.

11. Let X be an RV with PDF $f(x) = 1/(2\theta)$ if $-\theta \le x \le \theta$, and $= 0$ otherwise. Let $Y = 1/X^2$. Find the PDF of Y.

12. Let X be an RV of the continuous type, and let $Y = g(X)$ be defined as follows:

(a) $g(x) = 1$ if $x > 0$, and $= -1$ if $x \le 0$.

(b) $g(x) = b$ if $x \ge b$, $= x$ if $|x| < b$, and $= -b$ if $x \le -b$.

(c) $g(x) = x$ if $|x| \ge b$, and $= 0$ if $|x| < b$.

Find the distribution of Y in each case.