
INTRODUCTION

- 1 Introduction
- 2 Combinatorial Methods

- 3 Binomial Coefficients
- 4 The Theory in Practice

I Introduction

In recent years, the growth of statistics has made itself felt in almost every phase of human activity. Statistics no longer consists merely of the collection of data and their presentation in charts and tables; it is now considered to encompass the science of basing inferences on observed data and the entire problem of making decisions in the face of uncertainty. This covers considerable ground since uncertainties are met when we flip a coin, when a dietician experiments with food additives, when an actuary determines life insurance premiums, when a quality control engineer accepts or rejects manufactured products, when a teacher compares the abilities of students, when an economist forecasts trends, when a newspaper predicts an election, and even when a physicist describes quantum mechanics.

It would be presumptuous to say that statistics, in its present state of development, can handle all situations involving uncertainties, but new techniques are constantly being developed and modern statistics can, at least, provide the framework for looking at these situations in a logical and systematic fashion. In other words, statistics provides the models that are needed to study situations involving uncertainties, in the same way as calculus provides the models that are needed to describe, say, the concepts of Newtonian physics.

The beginnings of the mathematics of statistics may be found in mid-eighteenth-century studies in probability motivated by interest in games of chance. The theory thus developed for “heads or tails” or “red or black” soon found applications in situations where the outcomes were “boy or girl,” “life or death,” or “pass or fail,” and scholars began to apply probability theory to actuarial problems and some aspects of the social sciences. Later, probability and statistics were introduced into physics by L. Boltzmann, J. Gibbs, and J. Maxwell, and by this century they have found applications in all phases of human endeavor that in some way involve an element of uncertainty or risk. The names that are connected most prominently with the growth of mathematical statistics in the first half of the twentieth century are those of R. A. Fisher, J. Neyman, E. S. Pearson, and A. Wald. More recently, the work of R. Schlaifer, L. J. Savage, and others has given impetus to statistical theories based essentially on methods that date back to the eighteenth-century English clergyman Thomas Bayes.

Mathematical statistics is a recognized branch of mathematics, and it can be studied for its own sake by students of mathematics. Today, the theory of statistics is applied to engineering, physics and astronomy, quality assurance and reliability, drug development, public health and medicine, the design of agricultural or industrial experiments, experimental psychology, and so forth. Those wishing to participate

in such applications or to develop new applications will do well to understand the mathematical theory of statistics. For only through such an understanding can applications proceed without the serious mistakes that sometimes occur. The applications are illustrated by means of examples and a separate set of applied exercises, many of them involving the use of computers. To this end, we have added at the end of the chapter a discussion of how the theory of the chapter can be applied in practice.

We begin with a brief review of combinatorial methods and binomial coefficients.

2 Combinatorial Methods

In many problems of statistics we must list all the alternatives that are possible in a given situation, or at least determine how many different possibilities there are. In connection with the latter, we often use the following theorem, sometimes called the **basic principle of counting**, the **counting rule for compound events**, or the **rule for the multiplication of choices**.

THEOREM 1. If an operation consists of two steps, of which the first can be done in n_1 ways and for each of these the second can be done in n_2 ways, then the whole operation can be done in $n_1 \cdot n_2$ ways.

Here, “operation” stands for any kind of procedure, process, or method of selection.

To justify this theorem, let us define the ordered pair (x_i, y_j) to be the outcome that arises when the first step results in possibility x_i and the second step results in possibility y_j . Then, the set of all possible outcomes is composed of the following $n_1 \cdot n_2$ pairs:

$$\begin{aligned} &(x_1, y_1), (x_1, y_2), \dots, (x_1, y_{n_2}) \\ &(x_2, y_1), (x_2, y_2), \dots, (x_2, y_{n_2}) \\ &\quad \dots \\ &\quad \dots \\ &\quad \dots \\ &(x_{n_1}, y_1), (x_{n_1}, y_2), \dots, (x_{n_1}, y_{n_2}) \end{aligned}$$

EXAMPLE 1

Suppose that someone wants to go by bus, train, or plane on a week’s vacation to one of the five East North Central States. Find the number of different ways in which this can be done.

Solution

The particular state can be chosen in $n_1 = 5$ ways and the means of transportation can be chosen in $n_2 = 3$ ways. Therefore, the trip can be carried out in $5 \cdot 3 = 15$ possible ways. If an actual listing of all the possibilities is desirable, a **tree diagram** like that in Figure 1 provides a systematic approach. This diagram shows that there are $n_1 = 5$ branches (possibilities) for the number of states, and for each of these branches there are $n_2 = 3$ branches (possibilities) for the different means of transportation. It is apparent that the 15 possible ways of taking the vacation are represented by the 15 distinct paths along the branches of the tree.

Introduction

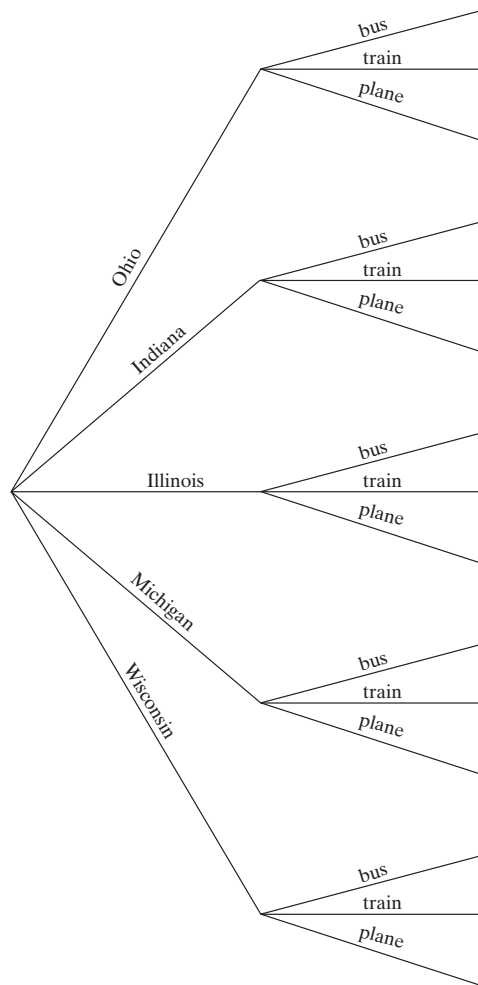


Figure 1. Tree diagram.

EXAMPLE 2

How many possible outcomes are there when we roll a pair of dice, one red and one green?

Solution

The red die can land in any one of six ways, and for each of these six ways the green die can also land in six ways. Therefore, the pair of dice can land in $6 \cdot 6 = 36$ ways.

Theorem 1 may be extended to cover situations where an operation consists of two or more steps. In this case, we have the following theorem.

THEOREM 2. If an operation consists of k steps, of which the first can be done in n_1 ways, for each of these the second step can be done in n_2 ways, for each of the first two the third step can be done in n_3 ways, and so forth, then the whole operation can be done in $n_1 \cdot n_2 \cdot \dots \cdot n_k$ ways.

EXAMPLE 3

A quality control inspector wishes to select a part for inspection from each of four different bins containing 4, 3, 5, and 4 parts, respectively. In how many different ways can she choose the four parts?

Solution

The total number of ways is $4 \cdot 3 \cdot 5 \cdot 4 = 240$.

EXAMPLE 4

In how many different ways can one answer all the questions of a true-false test consisting of 20 questions?

Solution

Altogether there are

$$2 \cdot 2 \cdot 2 \cdot 2 \cdot \dots \cdot 2 \cdot 2 = 2^{20} = 1,048,576$$

different ways in which one can answer all the questions; only one of these corresponds to the case where all the questions are correct and only one corresponds to the case where all the answers are wrong.

Frequently, we are interested in situations where the outcomes are the different ways in which a group of objects can be ordered or arranged. For instance, we might want to know in how many different ways the 24 members of a club can elect a president, a vice president, a treasurer, and a secretary, or we might want to know in how many different ways six persons can be seated around a table. Different arrangements like these are called **permutations**.

DEFINITION 1. PERMUTATIONS. A *permutation* is a distinct arrangement of n different elements of a set.

EXAMPLE 5

How many permutations are there of the letters a , b , and c ?

Solution

The possible arrangements are abc , acb , bac , bca , cab , and cba , so the number of distinct permutations is six. Using Theorem 2, we could have arrived at this answer without actually listing the different permutations. Since there are three choices to

Introduction

select a letter for the first position, then two for the second position, leaving only one letter for the third position, the total number of permutations is $3 \cdot 2 \cdot 1 = 6$.

Generalizing the argument used in the preceding example, we find that n distinct objects can be arranged in $n(n-1)(n-2) \cdots 3 \cdot 2 \cdot 1$ different ways. To simplify our notation, we represent this product by the symbol $n!$, which is read “ n factorial.” Thus, $1! = 1$, $2! = 2 \cdot 1 = 2$, $3! = 3 \cdot 2 \cdot 1 = 6$, $4! = 4 \cdot 3 \cdot 2 \cdot 1 = 24$, $5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$, and so on. Also, by definition we let $0! = 1$.

THEOREM 3. The number of permutations of n distinct objects is $n!$.

EXAMPLE 6

In how many different ways can the five starting players of a basketball team be introduced to the public?

Solution

There are $5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$ ways in which they can be introduced.

EXAMPLE 7

The number of permutations of the four letters a , b , c , and d is 24, but what is the number of permutations if we take only two of the four letters or, as it is usually put, if we take the four letters two at a time?

Solution

We have two positions to fill, with four choices for the first and then three choices for the second. Therefore, by Theorem 1, the number of permutations is $4 \cdot 3 = 12$.

Generalizing the argument that we used in the preceding example, we find that n distinct objects taken r at a time, for $r > 0$, can be arranged in $n(n-1) \cdots (n-r+1)$ ways. We denote this product by ${}_n P_r$, and we let ${}_n P_0 = 1$ by definition. Therefore, we can state the following theorem.

THEOREM 4. The number of permutations of n distinct objects taken r at a time is

$${}_n P_r = \frac{n!}{(n-r)!}$$

for $r = 0, 1, 2, \dots, n$.

Proof The formula ${}_n P_r = n(n-1) \cdots (n-r+1)$ cannot be used for $r = 0$, but we do have

$${}_n P_0 = \frac{n!}{(n-0)!} = 1$$

Introduction

For $r = 1, 2, \dots, n$, we have

$$\begin{aligned} {}_n P_r &= n(n-1)(n-2) \cdots (n-r+1) \\ &= \frac{n(n-1)(n-2) \cdots (n-r+1)(n-r)!}{(n-r)!} \\ &= \frac{n!}{(n-r)!} \end{aligned}$$

In problems concerning permutations, it is usually easier to proceed by using Theorem 2 as in Example 7, but the factorial formula of Theorem 4 is somewhat easier to remember. Many statistical software packages provide values of ${}_n P_r$ and other combinatorial quantities upon simple commands. Indeed, these quantities are also preprogrammed in many hand-held statistical (or scientific) calculators.

EXAMPLE 8

Four names are drawn from among the 24 members of a club for the offices of president, vice president, treasurer, and secretary. In how many different ways can this be done?

Solution

The number of permutations of 24 distinct objects taken four at a time is

$${}_{24} P_4 = \frac{24!}{20!} = 24 \cdot 23 \cdot 22 \cdot 21 = 255,024$$

EXAMPLE 9

In how many ways can a local chapter of the American Chemical Society schedule three speakers for three different meetings if they are all available on any of five possible dates?

Solution

Since we must choose three of the five dates and the order in which they are chosen (assigned to the three speakers) matters, we get

$${}_5 P_3 = \frac{5!}{2!} = \frac{120}{2} = 60$$

We might also argue that the first speaker can be scheduled in five ways, the second speaker in four ways, and the third speaker in three ways, so that the answer is $5 \cdot 4 \cdot 3 = 60$.

Permutations that occur when objects are arranged in a circle are called **circular permutations**. Two circular permutations are not considered different (and are counted only once) if corresponding objects in the two arrangements have the same objects to their left and to their right. For example, if four persons are playing bridge, we do not get a different permutation if everyone moves to the chair at his or her right.

EXAMPLE 10

How many circular permutations are there of four persons playing bridge?

Solution

If we arbitrarily consider the position of one of the four players as fixed, we can seat (arrange) the other three players in $3! = 6$ different ways. In other words, there are six different circular permutations.

Generalizing the argument used in the preceding example, we obtain the following theorem.

THEOREM 5. The number of permutations of n distinct objects arranged in a circle is $(n - 1)!$.

We have been assuming until now that the n objects from which we select r objects and form permutations are all distinct. Thus, the various formulas cannot be used, for example, to determine the number of ways in which we can arrange the letters in the word “book,” or the number of ways in which three copies of one novel and one copy each of four other novels can be arranged on a shelf.

EXAMPLE 11

How many different permutations are there of the letters in the word “book”?

Solution

If we distinguish for the moment between the two o 's by labeling them o_1 and o_2 , there are $4! = 24$ different permutations of the symbols b, o_1, o_2 , and k . However, if we drop the subscripts, then bo_1ko_2 and bo_2ko_1 , for instance, both yield $boko$, and since each pair of permutations with subscripts yields but one arrangement without subscripts, the total number of arrangements of the letters in the word “book” is $\frac{24}{2} = 12$.

EXAMPLE 12

In how many different ways can three copies of one novel and one copy each of four other novels be arranged on a shelf?

Solution

If we denote the three copies of the first novel by a_1, a_2 , and a_3 and the other four novels by b, c, d , and e , we find that *with subscripts* there are $7!$ different permutations of a_1, a_2, a_3, b, c, d , and e . However, since there are $3!$ permutations of a_1, a_2 , and a_3 that lead to the same permutation of a, a, a, b, c, d , and e , we find that there are only $\frac{7!}{3!} = 7 \cdot 6 \cdot 5 \cdot 4 = 840$ ways in which the seven books can be arranged on a shelf.

Generalizing the argument that we used in the two preceding examples, we obtain the following theorem.

THEOREM 6. The number of permutations of n objects of which n_1 are of one kind, n_2 are of a second kind, \dots , n_k are of a k th kind, and $n_1 + n_2 + \dots + n_k = n$ is

$$\frac{n!}{n_1! \cdot n_2! \cdot \dots \cdot n_k!}$$

EXAMPLE 13

In how many ways can two paintings by Monet, three paintings by Renoir, and two paintings by Degas be hung side by side on a museum wall if we do not distinguish between the paintings by the same artists?

Solution

Substituting $n = 7$, $n_1 = 2$, $n_2 = 3$, and $n_3 = 2$ into the formula of Theorem 6, we get

$$\frac{7!}{2! \cdot 3! \cdot 2!} = 210$$

There are many problems in which we are interested in determining the number of ways in which r objects can be selected from among n distinct objects *without regard to the order in which they are selected*.

DEFINITION 2. COMBINATIONS. A **combination** is a selection of r objects taken from n distinct objects without regard to the order of selection.

EXAMPLE 14

In how many different ways can a person gathering data for a market research organization select three of the 20 households living in a certain apartment complex?

Solution

If we care about the order in which the households are selected, the answer is

$${}_{20}P_3 = 20 \cdot 19 \cdot 18 = 6,840$$

but each set of three households would then be counted $3! = 6$ times. If we do not care about the order in which the households are selected, there are only $\frac{6,840}{6} = 1,140$ ways in which the person gathering the data can do his or her job.

Actually, “combination” means the same as “subset,” and when we ask for the number of combinations of r objects selected from a set of n distinct objects, we are simply asking for the total number of subsets of r objects that can be selected from a set of n distinct objects. In general, there are $r!$ permutations of the objects in a subset of r objects, so that the ${}_{n}P_r$ permutations of r objects selected from a set of n distinct objects contain each subset $r!$ times. Dividing ${}_{n}P_r$ by $r!$ and denoting the result by the symbol $\binom{n}{r}$, we thus have the following theorem.

THEOREM 7. The number of combinations of n distinct objects taken r at a time is

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$

for $r = 0, 1, 2, \dots, n$.

EXAMPLE 15

In how many different ways can six tosses of a coin yield two heads and four tails?

Solution

This question is the same as asking for the number of ways in which we can select the two tosses on which heads is to occur. Therefore, applying Theorem 7, we find that the answer is

$$\binom{6}{2} = \frac{6!}{2! \cdot 4!} = 15$$

This result could also have been obtained by the rather tedious process of enumerating the various possibilities, HHTTTT, TTHTHT, HTHTTT, ..., where H stands for head and T for tail.

EXAMPLE 16

How many different committees of two chemists and one physicist can be formed from the four chemists and three physicists on the faculty of a small college?

Solution

Since two of four chemists can be selected in $\binom{4}{2} = \frac{4!}{2! \cdot 2!} = 6$ ways and one of three physicists can be selected in $\binom{3}{1} = \frac{3!}{1! \cdot 2!} = 3$ ways, Theorem 1 shows that the number of committees is $6 \cdot 3 = 18$.

A combination of r objects selected from a set of n distinct objects may be considered a **partition** of the n objects into two subsets containing, respectively, the r objects that are selected and the $n - r$ objects that are left. Often, we are concerned with the more general problem of partitioning a set of n distinct objects into k subsets, which requires that each of the n objects must belong to one and only one of the subsets.[†] The order of the objects within a subset is of no importance.

EXAMPLE 17

In how many ways can a set of four objects be partitioned into three subsets containing, respectively, two, one, and one of the objects?

[†]Symbolically, the subsets A_1, A_2, \dots, A_k constitute a partition of set A if $A_1 \cup A_2 \cup \dots \cup A_k = A$ and $A_i \cap A_j = \emptyset$ for all $i \neq j$.

Solution

Denoting the four objects by $a, b, c,$ and $d,$ we find by enumeration that there are the following 12 possibilities:

$$\begin{array}{cccc} ab|c|d & ab|d|c & ac|b|d & ac|d|b \\ ad|b|c & ad|c|b & bc|a|d & bc|d|a \\ bd|a|c & bd|c|a & cd|a|b & cd|b|a \end{array}$$

The number of partitions for this example is denoted by the symbol

$$\binom{4}{2,1,1} = 12$$

where the number at the top represents the total number of objects and the numbers at the bottom represent the number of objects going into each subset.

Had we not wanted to enumerate all the possibilities in the preceding example, we could have argued that the two objects going into the first subset can be chosen in $\binom{4}{2} = 6$ ways, the object going into the second subset can then be chosen in $\binom{2}{1} = 2$ ways, and the object going into the third subset can then be chosen in $\binom{1}{1} = 1$ way. Thus, by Theorem 2 there are $6 \cdot 2 \cdot 1 = 12$ partitions. Generalizing this argument, we have the following theorem.

THEOREM 8. The number of ways in which a set of n distinct objects can be partitioned into k subsets with n_1 objects in the first subset, n_2 objects in the second subset, $\dots,$ and n_k objects in the k th subset is

$$\binom{n}{n_1, n_2, \dots, n_k} = \frac{n!}{n_1! \cdot n_2! \cdot \dots \cdot n_k!}$$

Proof Since the n_1 objects going into the first subset can be chosen in $\binom{n}{n_1}$ ways, the n_2 objects going into the second subset can then be chosen in $\binom{n-n_1}{n_2}$ ways, the n_3 objects going into the third subset can then be chosen in $\binom{n-n_1-n_2}{n_3}$ ways, and so forth, it follows by Theorem 2 that the total number of partitions is

$$\begin{aligned} \binom{n}{n_1, n_2, \dots, n_k} &= \binom{n}{n_1} \cdot \binom{n-n_1}{n_2} \cdot \dots \cdot \binom{n-n_1-n_2-\dots-n_{k-1}}{n_k} \\ &= \frac{n!}{n_1! \cdot (n-n_1)!} \cdot \frac{(n-n_1)!}{n_2! \cdot (n-n_1-n_2)!} \\ &\quad \cdot \dots \cdot \frac{(n-n_1-n_2-\dots-n_{k-1})!}{n_k! \cdot 0!} \\ &= \frac{n!}{n_1! \cdot n_2! \cdot \dots \cdot n_k!} \end{aligned}$$

EXAMPLE 18

In how many ways can seven businessmen attending a convention be assigned to one triple and two double hotel rooms?

Solution

Substituting $n = 7$, $n_1 = 3$, $n_2 = 2$, and $n_3 = 2$ into the formula of Theorem 8, we get

$$\binom{7}{3, 2, 2} = \frac{7!}{3! \cdot 2! \cdot 2!} = 210$$

3 Binomial Coefficients

If n is a positive integer and we multiply out $(x + y)^n$ term by term, each term will be the product of x 's and y 's, with an x or a y coming from each of the n factors $x + y$. For instance, the expansion

$$\begin{aligned} (x + y)^3 &= (x + y)(x + y)(x + y) \\ &= x \cdot x \cdot x + x \cdot x \cdot y + x \cdot y \cdot x + x \cdot y \cdot y \\ &\quad + y \cdot x \cdot x + y \cdot x \cdot y + y \cdot y \cdot x + y \cdot y \cdot y \\ &= x^3 + 3x^2y + 3xy^2 + y^3 \end{aligned}$$

yields terms of the form x^3 , x^2y , xy^2 , and y^3 . Their coefficients are 1, 3, 3, and 1, and the coefficient of xy^2 , for example, is $\binom{3}{2} = 3$, the number of ways in which we can choose the two factors providing the y 's. Similarly, the coefficient of x^2y is $\binom{3}{1} = 3$, the number of ways in which we can choose the one factor providing the y , and the coefficients of x^3 and y^3 are $\binom{3}{0} = 1$ and $\binom{3}{3} = 1$.

More generally, if n is a positive integer and we multiply out $(x + y)^n$ term by term, the coefficient of $x^{n-r}y^r$ is $\binom{n}{r}$, the number of ways in which we can choose the r factors providing the y 's. Accordingly, we refer to $\binom{n}{r}$ as a **binomial coefficient**. Values of the binomial coefficients for $n = 0, 1, \dots, 20$ and $r = 0, 1, \dots, 10$ are given in table Factorials and Binomial Coefficients of "Statistical Tables." We can now state the following theorem.

THEOREM 9.

$$(x + y)^n = \sum_{r=0}^n \binom{n}{r} x^{n-r} y^r \quad \text{for any positive integer } n$$

DEFINITION 3. BINOMIAL COEFFICIENTS. The coefficient of $x^{n-r}y^r$ in the binomial expansion of $(x+y)^n$ is called the **binomial coefficient** $\binom{n}{r}$.

The calculation of binomial coefficients can often be simplified by making use of the three theorems that follow.

THEOREM 10. For any positive integers n and $r = 0, 1, 2, \dots, n$,

$$\binom{n}{r} = \binom{n}{n-r}$$

Proof We might argue that when we select a subset of r objects from a set of n distinct objects, we leave a subset of $n-r$ objects; hence, there are as many ways of selecting r objects as there are ways of leaving (or selecting) $n-r$ objects. To prove the theorem algebraically, we write

$$\begin{aligned} \binom{n}{n-r} &= \frac{n!}{(n-r)![n-(n-r)]!} = \frac{n!}{(n-r)!r!} \\ &= \frac{n!}{r!(n-r)!} = \binom{n}{r} \end{aligned}$$

Theorem 10 implies that if we calculate the binomial coefficients for $r = 0, 1, \dots, \frac{n}{2}$ when n is even and for $r = 0, 1, \dots, \frac{n-1}{2}$ when n is odd, the remaining binomial coefficients can be obtained by making use of the theorem.

EXAMPLE 19

Given $\binom{4}{0} = 1$, $\binom{4}{1} = 4$, and $\binom{4}{2} = 6$, find $\binom{4}{3}$ and $\binom{4}{4}$.

Solution

$$\binom{4}{3} = \binom{4}{4-3} = \binom{4}{1} = 4 \text{ and } \binom{4}{4} = \binom{4}{4-4} = \binom{4}{0} = 1$$

EXAMPLE 20

Given $\binom{5}{0} = 1$, $\binom{5}{1} = 5$, and $\binom{5}{2} = 10$, find $\binom{5}{3}$, $\binom{5}{4}$, and $\binom{5}{5}$.

Solution

$$\begin{aligned} \binom{5}{3} &= \binom{5}{5-3} = \binom{5}{2} = 10, \binom{5}{4} = \binom{5}{5-4} = \binom{5}{1} = 5, \text{ and} \\ \binom{5}{5} &= \binom{5}{5-5} = \binom{5}{0} = 1 \end{aligned}$$

It is precisely in this fashion that Theorem 10 may have to be used in connection with table Factorials and Binomial Coefficients of "Statistical Tables."

EXAMPLE 21

Find $\binom{20}{12}$ and $\binom{17}{10}$.

Solution

Since $\binom{20}{12}$ is not given in the table, we make use of the fact that $\binom{20}{12} = \binom{20}{8}$, look up $\binom{20}{8}$, and get $\binom{20}{12} = 125,970$. Similarly, to find $\binom{17}{10}$, we make use of the fact that $\binom{17}{10} = \binom{17}{7}$, look up $\binom{17}{7}$, and get $\binom{17}{10} = 19,448$.

THEOREM 11. For any positive integer n and $r = 1, 2, \dots, n-1$,

$$\binom{n}{r} = \binom{n-1}{r} + \binom{n-1}{r-1}$$

Proof Substituting $x = 1$ into $(x+y)^n$, let us write $(1+y)^n = (1+y)(1+y)^{n-1} = (1+y)^{n-1} + y(1+y)^{n-1}$ and equate the coefficient of y^r in $(1+y)^n$ with that in $(1+y)^{n-1} + y(1+y)^{n-1}$. Since the coefficient of y^r in $(1+y)^n$ is $\binom{n}{r}$ and the coefficient of y^r in $(1+y)^{n-1} + y(1+y)^{n-1}$ is the sum of the coefficient of y^r in $(1+y)^{n-1}$, that is, $\binom{n-1}{r}$, and the coefficient of y^{r-1} in $(1+y)^{n-1}$, that is, $\binom{n-1}{r-1}$, we obtain

$$\binom{n}{r} = \binom{n-1}{r} + \binom{n-1}{r-1}$$

which completes the proof.

Alternatively, take any one of the n objects. If it is not to be included among the r objects, there are $\binom{n-1}{r}$ ways of selecting the r objects; if it is to be included, there are $\binom{n-1}{r-1}$ ways of selecting the other $r-1$ objects. Therefore, there are $\binom{n-1}{r} + \binom{n-1}{r-1}$ ways of selecting the r objects, that is,

$$\binom{n}{r} = \binom{n-1}{r} + \binom{n-1}{r-1}$$

Theorem 11 can also be proved by expressing the binomial coefficients on both sides of the equation in terms of factorials and then proceeding algebraically, but we shall leave this to the reader in Exercise 12.

An important application of Theorem 11 is a construct known as **Pascal's triangle**. When no table is available, it is sometimes convenient to determine binomial coefficients by means of a simple construction. Applying Theorem 11, we can generate Pascal's triangle as follows:

Introduction

$$\begin{array}{ccccccc}
 & & & & 1 & & & & \\
 & & & & 1 & & 1 & & \\
 & & & 1 & 2 & 1 & & & \\
 & & 1 & 3 & 3 & 1 & & & \\
 & 1 & 4 & 6 & 4 & 1 & & & \\
 1 & 5 & 10 & 10 & 5 & 1 & & & \\
 \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots
 \end{array}$$

In this triangle, the first and last entries of each row are the numeral “1” each other entry in any given row is obtained by adding the two entries in the preceding row immediately to its left and to its right.

To state the third theorem about binomial coefficients, let us make the following definition: $\binom{n}{r} = 0$ whenever n is a positive integer and r is a positive integer greater than n . (Clearly, there is no way in which we can select a subset that contains more elements than the whole set itself.)

THEOREM 12.

$$\sum_{r=0}^k \binom{m}{r} \binom{n}{k-r} = \binom{m+n}{k}$$

Proof Using the same technique as in the proof of Theorem 11, let us prove this theorem by equating the coefficients of y^k in the expressions on both sides of the equation

$$(1+y)^{m+n} = (1+y)^m(1+y)^n$$

The coefficient of y^k in $(1+y)^{m+n}$ is $\binom{m+n}{k}$, and the coefficient of y^k in

$$\begin{aligned}
 (1+y)^m(1+y)^n &= \left[\binom{m}{0} + \binom{m}{1}y + \dots + \binom{m}{m}y^m \right] \\
 &\quad \times \left[\binom{n}{0} + \binom{n}{1}y + \dots + \binom{n}{n}y^n \right]
 \end{aligned}$$

is the sum of the products that we obtain by multiplying the constant term of the first factor by the coefficient of y^k in the second factor, the coefficient of y in the first factor by the coefficient of y^{k-1} in the second factor, . . . , and the coefficient of y^k in the first factor by the constant term of the second factor. Thus, the coefficient of y^k in $(1+y)^m(1+y)^n$ is

$$\begin{aligned}
 &\binom{m}{0} \binom{n}{k} + \binom{m}{1} \binom{n}{k-1} + \binom{m}{2} \binom{n}{k-2} + \dots + \binom{m}{k} \binom{n}{0} \\
 &= \sum_{r=0}^k \binom{m}{r} \binom{n}{k-r}
 \end{aligned}$$

and this completes the proof.

EXAMPLE 22

Verify Theorem 12 numerically for $m = 2$, $n = 3$, and $k = 4$.

Solution

Substituting these values, we get

$$\binom{2}{0} \binom{3}{4} + \binom{2}{1} \binom{3}{3} + \binom{2}{2} \binom{3}{2} + \binom{2}{3} \binom{3}{1} + \binom{2}{4} \binom{3}{0} = \binom{5}{4}$$

and since $\binom{3}{4}$, $\binom{2}{3}$, and $\binom{2}{4}$ equal 0 according to the definition on the previous page, the equation reduces to

$$\binom{2}{1} \binom{3}{3} + \binom{2}{2} \binom{3}{2} = \binom{5}{4}$$

which checks, since $2 \cdot 1 + 1 \cdot 3 = 5$.

Using Theorem 8, we can extend our discussion to **multinomial coefficients**, that is, to the coefficients that arise in the expansion of $(x_1 + x_2 + \cdots + x_k)^n$. The multinomial coefficient of the term $x_1^{r_1} \cdot x_2^{r_2} \cdot \cdots \cdot x_k^{r_k}$ in the expansion of $(x_1 + x_2 + \cdots + x_k)^n$ is

$$\binom{n}{r_1, r_2, \dots, r_k} = \frac{n!}{r_1! \cdot r_2! \cdot \cdots \cdot r_k!}$$

EXAMPLE 23

What is the coefficient of $x_1^3 x_2 x_3^2$ in the expansion of $(x_1 + x_2 + x_3)^6$?

Solution

Substituting $n = 6$, $r_1 = 3$, $r_2 = 1$, and $r_3 = 2$ into the preceding formula, we get

$$\frac{6!}{3! \cdot 1! \cdot 2!} = 60$$

Exercises

1. An operation consists of two steps, of which the first can be made in n_1 ways. If the first step is made in the i th way, the second step can be made in n_{2i} ways.[†]

(a) Use a tree diagram to find a formula for the total number of ways in which the total operation can be made.

(b) A student can study 0, 1, 2, or 3 hours for a history test on any given day. Use the formula obtained in part (a) to verify that there are 13 ways in which the student can study at most 4 hours for the test on two consecutive days.

2. With reference to Exercise 1, verify that if n_{2i} equals the constant n_2 , the formula obtained in part (a) reduces to that of Theorem 1.

3. With reference to Exercise 1, suppose that there is a third step, and if the first step is made in the i th way and the second step in the j th way, the third step can be made in n_{3ij} ways.

(a) Use a tree diagram to verify that the whole operation can be made in

[†]The first subscript denotes the row to which a particular element belongs, and the second subscript denotes the column.

$$\sum_{i=1}^{n_1} \sum_{j=1}^{n_{2i}} n_{3ij}$$

different ways.

(b) With reference to part (b) of Exercise 1, use the formula of part (a) to verify that there are 32 ways in which the student can study at most 4 hours for the test on three consecutive days.

4. Show that if n_{2i} equals the constant n_2 and n_{3ij} equals the constant n_3 , the formula of part (a) of Exercise 3 reduces to that of Theorem 2.

5. In a two-team basketball play-off, the winner is the first team to win m games.

(a) Counting separately the number of play-offs requiring $m, m+1, \dots$, and $2m-1$ games, show that the total number of different outcomes (sequences of wins and losses by one of the teams) is

$$2 \left[\binom{m-1}{m-1} + \binom{m}{m-1} + \dots + \binom{2m-2}{m-1} \right]$$

(b) How many different outcomes are there in a “2 out of 3” play-off, a “3 out of 5” play-off, and a “4 out of 7” play-off?

6. When n is large, $n!$ can be approximated by means of the expression

$$\sqrt{2\pi n} \left(\frac{n}{e} \right)^n$$

called **Stirling’s formula**, where e is the base of natural logarithms. (A derivation of this formula may be found in the book by W. Feller cited among the references at the end of this chapter.)

(a) Use Stirling’s formula to obtain approximations for $10!$ and $12!$, and find the percentage errors of these approximations by comparing them with the exact values given in table Factorials and Binomial Coefficients of “Statistical Tables.”

(b) Use Stirling’s formula to obtain an approximation for the number of 13-card bridge hands that can be dealt with an ordinary deck of 52 playing cards.

7. Using Stirling’s formula (see Exercise 6) to approximate $2n!$ and $n!$, show that

$$\frac{\binom{2n}{n} \sqrt{\pi n}}{2^{2n}} \approx 1$$

8. In some problems of **occupancy theory** we are concerned with the number of ways in which certain *distinguishable* objects can be distributed among individuals, urns, boxes, or cells. Find an expression for the number of ways in which r *distinguishable* objects can be distributed among n cells, and use it to find the number of ways in

which three different books can be distributed among the 12 students in an English literature class.

9. In some problems of occupancy theory we are concerned with the number of ways in which certain *indistinguishable* objects can be distributed among individuals, urns, boxes, or cells. Find an expression for the number of ways in which r *indistinguishable* objects can be distributed among n cells, and use it to find the number of ways in which a baker can sell five (indistinguishable) loaves of bread to three customers. (*Hint*: We might argue that L|LLL|L represents the case where the three customers buy one loaf, three loaves, and one loaf, respectively, and that LLLL|L represents the case where the three customers buy four loaves, none of the loaves, and one loaf. Thus, we must look for the number of ways in which we can arrange the five L’s and the two vertical bars.)

10. In some problems of occupancy theory we are concerned with the number of ways in which certain *indistinguishable* objects can be distributed among individuals, urns, boxes, or cells with at least one in each cell. Find an expression for the number of ways in which r *indistinguishable* objects can be distributed among n cells with at least one in each cell, and rework the numerical part of Exercise 9 with each of the three customers getting at least one loaf of bread.

11. Construct the seventh and eighth rows of Pascal’s triangle and write the binomial expansions of $(x+y)^6$ and $(x+y)^7$.

12. Prove Theorem 11 by expressing all the binomial coefficients in terms of factorials and then simplifying algebraically.

13. Expressing the binomial coefficients in terms of factorials and simplifying algebraically, show that

$$(a) \binom{n}{r} = \frac{n-r+1}{r} \cdot \binom{n}{r-1};$$

$$(b) \binom{n}{r} = \frac{n}{n-r} \cdot \binom{n-1}{r};$$

$$(c) n \binom{n-1}{r} = (r+1) \binom{n}{r+1}.$$

14. Substituting appropriate values for x and y into the formula of Theorem 9, show that

$$(a) \sum_{r=0}^n \binom{n}{r} = 2^n;$$

$$(b) \sum_{r=0}^n (-1)^r \binom{n}{r} = 0;$$

$$(c) \sum_{r=0}^n \binom{n}{r} (a-1)^r = a^n.$$

15. Repeatedly applying Theorem 11, show that

$$\binom{n}{r} = \sum_{i=1}^{r+1} \binom{n-i}{r-i+1}$$

16. Use Theorem 12 to show that

$$\sum_{r=0}^n \binom{n}{r}^2 = \binom{2n}{n}$$

17. Show that $\sum_{r=0}^n r \binom{n}{r} = n2^{n-1}$ by setting $x = 1$ in Theorem 9, then differentiating the expressions on both sides with respect to y , and finally substituting $y = 1$.

18. Rework Exercise 17 by making use of part (a) of Exercise 14 and part (c) of Exercise 13.

19. If n is not a positive integer or zero, the binomial expansion of $(1+y)^n$ yields, for $-1 < y < 1$, the infinite series

$$1 + \binom{n}{1}y + \binom{n}{2}y^2 + \binom{n}{3}y^3 + \cdots + \binom{n}{r}y^r + \cdots$$

where $\binom{n}{r} = \frac{n(n-1) \cdots (n-r+1)}{r!}$ for $r = 1, 2, 3, \dots$

Use this **generalized definition of binomial coefficients** to evaluate

(a) $\binom{\frac{1}{2}}{4}$ and $\binom{-3}{3}$;

(b) $\sqrt{5}$ writing $\sqrt{5} = 2(1 + \frac{1}{4})^{1/2}$ and using the first four terms of the binomial expansion of $(1 + \frac{1}{4})^{1/2}$.

20. With reference to the generalized definition of binomial coefficients in Exercise 19, show that

(a) $\binom{-1}{r} = (-1)^r$;

(b) $\binom{-n}{r} = (-1)^r \binom{n+r-1}{r}$ for $n > 0$.

21. Find the coefficient of $x^2y^3z^3$ in the expansion of $(x+y+z)^8$.

22. Find the coefficient of $x^3y^2z^3w$ in the expansion of $(2x+3y-4z+w)^9$.

23. Show that

$$\begin{aligned} \binom{n}{n_1, n_2, \dots, n_k} &= \binom{n-1}{n_1-1, n_2, \dots, n_k} \\ &+ \binom{n-1}{n_1, n_2-1, \dots, n_k} + \cdots \\ &+ \binom{n-1}{n_1, n_2, \dots, n_k-1} \end{aligned}$$

by expressing all these multinomial coefficients in terms of factorials and simplifying algebraically.

4 The Theory in Practice

Applications of the preceding theory of combinatorial methods and binomial coefficients are quite straightforward, and a variety of them have been given in Sections 2 and 3. The following examples illustrate further applications of this theory.

EXAMPLE 24

An assembler of electronic equipment has 20 integrated-circuit chips on her table, and she must solder three of them as part of a larger component. In how many ways can she choose the three chips for assembly?

Solution

Using Theorem 6, we obtain the result

$${}_{20}P_3 = 20!/17! = 20 \cdot 19 \cdot 18 = 6,840$$

EXAMPLE 25

A lot of manufactured goods, presented for sampling inspection, contains 16 units. In how many ways can 4 of the 16 units be selected for inspection?

Solution

According to Theorem 7,

$$\binom{16}{4} = 16!/4!12! = 16 \cdot 15 \cdot 14 \cdot 13/4 \cdot 3 \cdot 2 \cdot 1 = 1,092 \text{ ways}$$

Applied Exercises**SECS. 1–4**

- 24.** A thermostat will call for heat 0, 1, or 2 times a night. Construct a tree diagram to show that there are 10 different ways that it can turn on the furnace for a total of 6 times over 4 nights.
- 25.** On August 31 there are five wild-card terms in the American League that can make it to the play-offs, and only two will win spots. Draw a tree diagram which shows the various possible play-off wild-card teams.
- 26.** There are four routes, A , B , C , and D , between a person's home and the place where he works, but route B is one-way, so he cannot take it on the way to work, and route C is one-way, so he cannot take it on the way home.
(a) Draw a tree diagram showing the various ways the person can go to and from work.
(b) Draw a tree diagram showing the various ways he can go to and from work without taking the same route both ways.
- 27.** A person with \$2 in her pocket bets \$1, even money, on the flip of a coin, and she continues to bet \$1 as long as she has any money. Draw a tree diagram to show the various things that can happen during the first four flips of the coin. After the fourth flip of the coin, in how many of the cases will she be
(a) exactly even;
(b) exactly \$2 ahead?
- 28.** The pro at a golf course stocks two identical sets of women's clubs, reordering at the end of each day (for delivery early the next morning) if and only if he has sold them both. Construct a tree diagram to show that if he starts on a Monday with two sets of the clubs, there are altogether eight different ways in which he can make sales on the first two days of that week.
- 29.** Suppose that in a baseball World Series (in which the winner is the first team to win four games) the National League champion leads the American League champion three games to two. Construct a tree diagram to show the number of ways in which these teams may win or lose the remaining game or games.
- 30.** If the NCAA has applications from six universities for hosting its intercollegiate tennis championships in two consecutive years, in how many ways can they select the hosts for these championships
(a) if they are not both to be held at the same university;
(b) if they may both be held at the same university?
- 31.** Counting the number of outcomes in games of chance has been a popular pastime for many centuries. This was of interest not only because of the gambling that was involved, but also because the outcomes of games of chance were often interpreted as divine intent. Thus, it was just about a thousand years ago that a bishop in what is now Belgium determined that there are 56 different ways in which three dice can fall *provided one is interested only in the overall result and not in which die does what*. He assigned a virtue to each of these possibilities and each sinner had to concentrate for some time on the virtue that corresponded to his cast of the dice.
(a) Find the number of ways in which three dice can all come up with the same number of points.
(b) Find the number of ways in which two of the three dice can come up with the same number of points, while the third comes up with a different number of points.
(c) Find the number of ways in which all three of the dice can come up with a different number of points.
(d) Use the results of parts (a), (b), and (c) to verify the bishop's calculations that there are altogether 56 possibilities.
- 32.** In a primary election, there are four candidates for mayor, five candidates for city treasurer, and two candidates for county attorney.
(a) In how many ways can a voter mark his ballot for all three of these offices?
(b) In how many ways can a person vote if he exercises his option of not voting for a candidate for any or all of these offices?
- 33.** The five finalists in the Miss Universe contest are Miss Argentina, Miss Belgium, Miss U.S.A., Miss Japan, and Miss Norway. In how many ways can the judges choose
(a) the winner and the first runner-up;
(b) the winner, the first runner-up, and the second runner-up?

Introduction

- 34.** A multiple-choice test consists of 15 questions, each permitting a choice of three alternatives. In how many different ways can a student check off her answers to these questions?
- 35.** Determine the number of ways in which a distributor can choose 2 of 15 warehouses to ship a large order.
- 36.** The price of a European tour includes four stopovers to be selected from among 10 cities. In how many different ways can one plan such a tour
(a) if the order of the stopovers matters;
(b) if the order of the stopovers does not matter?
- 37.** A carton of 15 light bulbs contains one that is defective. In how many ways can an inspector choose 3 of the bulbs and
(a) get the one that is defective;
(b) not get the one that is defective?
- 38.** In how many ways can a television director schedule a sponsor's six different commercials during the six time slots allocated to commercials during a two-hour program?
- 39.** In how many ways can the television director of Exercise 38 fill the six time slots for commercials if there are three different sponsors and the commercial for each is to be shown twice?
- 40.** In how many ways can five persons line up to get on a bus? In how many ways can they line up if two of the persons refuse to follow each other?
- 41.** In how many ways can eight persons form a circle for a folk dance?
- 42.** How many permutations are there of the letters in the word
(a) "great";
(b) "greet"?
- 43.** How many distinct permutations are there of the letters in the word "statistics"? How many of these begin and end with the letter *s*?
- 44.** A college team plays 10 football games during a season. In how many ways can it end the season with five wins, four losses, and one tie?
- 45.** If eight persons are having dinner together, in how many different ways can three order chicken, four order steak, and one order lobster?
- 46.** In Example 4 we showed that a true–false test consisting of 20 questions can be marked in 1,048,576 different ways. In how many ways can each question be marked true or false so that
(a) 7 are right and 13 are wrong;
(b) 10 are right and 10 are wrong;
(c) at least 17 are right?
- 47.** Among the seven nominees for two vacancies on a city council are three men and four women. In how many ways can these vacancies be filled
(a) with any two of the seven nominees;
(b) with any two of the four women;
(c) with one of the men and one of the women?
- 48.** A shipment of 10 television sets includes three that are defective. In how many ways can a hotel purchase four of these sets and receive at least two of the defective sets?
- 49.** Ms. Jones has four skirts, seven blouses, and three sweaters. In how many ways can she choose two of the skirts, three of the blouses, and one of the sweaters to take along on a trip?
- 50.** How many different bridge hands are possible containing five spades, three diamonds, three clubs, and two hearts?
- 51.** Find the number of ways in which one A, three B's, two C's, and one F can be distributed among seven students taking a course in statistics.
- 52.** An art collector, who owns 10 paintings by famous artists, is preparing her will. In how many different ways can she leave these paintings to her three heirs?
- 53.** A baseball fan has a pair of tickets for six different home games of the Chicago Cubs. If he has five friends who like baseball, in how many different ways can he take one of them along to each of the six games?
- 54.** At the end of the day, a bakery gives everything that is unsold to food banks for the needy. If it has 12 apple pies left at the end of a given day, in how many different ways can it distribute these pies among six food banks for the needy?
- 55.** With reference to Exercise 54, in how many different ways can the bakery distribute the 12 apple pies if each of the six food banks is to receive at least one pie?
- 56.** On a Friday morning, the pro shop of a tennis club has 14 identical cans of tennis balls. If they are all sold by Sunday night and we are interested only in how many were sold on each day, in how many different ways could the tennis balls have been sold on Friday, Saturday, and Sunday?
- 57.** Rework Exercise 56 given that at least two of the cans of tennis balls were sold on each of the three days.

References

- Among the books on the history of statistics there are
- WALKER, H. M., *Studies in the History of Statistical Method*. Baltimore: The Williams & Wilkins Company, 1929,
- WESTERGAARD, H., *Contributions to the History of Statistics*. London: P. S. King & Son, 1932,
- and the more recent publications
- KENDALL, M. G., and PLACKETT, R. L., eds., *Studies in the History of Statistics and Probability*, Vol. II. New York: Macmillan Publishing Co., Inc., 1977,
- PEARSON, E. S., and KENDALL, M. G., eds., *Studies in the History of Statistics and Probability*. Darien, Conn.: Hafner Publishing Co., Inc., 1970,
- PORTER, T. M., *The Rise of Statistical Thinking, 1820–1900*. Princeton, N.J.: Princeton University Press, 1986,
- and
- STIGLER, S. M., *The History of Statistics*. Cambridge, Mass.: Harvard University Press, 1986.
- A wealth of material on combinatorial methods can be found in
- COHEN, D. A., *Basic Techniques of Combinatorial Theory*. New York: John Wiley & Sons, Inc., 1978,
- EISEN, M., *Elementary Combinatorial Analysis*. New York: Gordon and Breach, Science Publishers, Inc., 1970,
- FELLER, W., *An Introduction to Probability Theory and Its Applications*, Vol. I, 3rd ed. New York: John Wiley & Sons, Inc., 1968,
- NIVEN, J., *Mathematics of Choice*. New York: Random House, Inc., 1965,
- ROBERTS, F. S., *Applied Combinatorics*. Upper Saddle River, N.J.: Prentice Hall, 1984,
- and
- WHITWORTH, W. A., *Choice and Chance*, 5th ed. New York: Hafner Publishing Co., Inc., 1959, which has become a classic in this field.
- More advanced treatments may be found in
- BECKENBACH, E. F., ed., *Applied Combinatorial Mathematics*. New York: John Wiley & Sons, Inc., 1964,
- DAVID, F. N., and BARTON, D. E., *Combinatorial Chance*. New York: Hafner Publishing Co., Inc., 1962,
- and
- RIORDAN, J., *An Introduction to Combinatorial Analysis*. New York: John Wiley & Sons, Inc., 1958.

Answers to Odd-Numbered Exercises

- 1 (a) $\sum_{i=1}^n n2n_i$.
- 5 (b) 6, 20, and 70.
- 9 $\binom{r+n-1}{r}$ and 21.
- 11 Seventh row: 1, 6, 15, 20, 15, 6, 1; Eighth row: 1, 7, 21, 35, 35, 21, 7, 1.
- 19 (a) $\frac{-15}{384}$ and -10 ; (b) 2.230.
- 21 560.
- 27 (a) 5; (b) 4.
- 31 (a) 6; (b) 30; (c) 20; (d) 56.
- 33 (a) 20; (b) 60.
- 35 (a) 105.
- 37 (a) 91; (b) 364.
- 39 90.
- 41 5040.
- 43 50,400 and 3360.
- 45 280.
- 47 (a) 21; (b) 6; (c) 12.
- 49 630.
- 51 420.
- 53 15,625.
- 55 462.
- 57 45.

PROBABILITY

- | | |
|-------------------------------|---------------------------|
| 1 Introduction | 6 Conditional Probability |
| 2 Sample Spaces | 7 Independent Events |
| 3 Events | 8 Bayes' Theorem |
| 4 The Probability of an Event | 9 The Theory in Practice |
| 5 Some Rules of Probability | |

I Introduction

Historically, the oldest way of defining probabilities, the **classical probability concept**, applies when *all possible outcomes are equally likely*, as is presumably the case in most games of chance. We can then say that *if there are N equally likely possibilities, of which one must occur and n are regarded as favorable, or as a "success," then the probability of a "success" is given by the ratio $\frac{n}{N}$.*

EXAMPLE 1

What is the probability of drawing an ace from an ordinary deck of 52 playing cards?

Solution

Since there are $n = 4$ aces among the $N = 52$ cards, the probability of drawing an ace is $\frac{4}{52} = \frac{1}{13}$. (It is assumed, of course, that each card has the same chance of being drawn.)

Although equally likely possibilities are found mostly in games of chance, the classical probability concept applies also in a great variety of situations where gambling devices are used to make random selections—when office space is assigned to teaching assistants by lot, when some of the families in a township are chosen in such a way that each one has the same chance of being included in a sample study, when machine parts are chosen for inspection so that each part produced has the same chance of being selected, and so forth.

A major shortcoming of the classical probability concept is its limited applicability, for there are many situations in which the possibilities that arise cannot all be regarded as equally likely. This would be the case, for instance, if we are concerned with the question whether it will rain on a given day, if we are concerned with the outcome of an election, or if we are concerned with a person's recovery from a disease.

Among the various probability concepts, most widely held is the **frequency interpretation**, according to which *the probability of an event (outcome or happening)* is the proportion of the time that events of the same kind will occur in the long run. If we say that the probability is 0.84 that a jet from Los Angeles to San Francisco will arrive on time, we mean (in accordance with the frequency interpretation) that such flights arrive on time 84 percent of the time. Similarly, if the weather bureau

From Chapter 2 of *John E. Freund's Mathematical Statistics with Applications*, Eighth Edition. Irwin Miller, Marylees Miller. Copyright © 2014 by Pearson Education, Inc. All rights reserved.

predicts that there is a 30 percent chance for rain (that is, a probability of 0.30), this means that under the same weather conditions it will rain 30 percent of the time. More generally, we say that an event has a probability of, say, 0.90, in the same sense in which we might say that our car will start in cold weather 90 percent of the time. We cannot guarantee what will happen on any particular occasion—the car may start and then it may not—but if we kept records over a long period of time, we should find that the proportion of “successes” is very close to 0.90.

The approach to probability that we shall use in this chapter is the **axiomatic approach**, in which probabilities are defined as “mathematical objects” that behave according to certain well-defined rules. Then, any one of the preceding probability concepts, or interpretations, can be used in applications as long as it is consistent with these rules.

2 Sample Spaces

Since all probabilities pertain to the occurrence or nonoccurrence of events, let us explain first what we mean here by *event* and by the related terms *experiment*, *outcome*, and *sample space*.

It is customary in statistics to refer to any process of observation or measurement as an **experiment**. In this sense, an experiment may consist of the simple process of checking whether a switch is turned on or off; it may consist of counting the imperfections in a piece of cloth; or it may consist of the very complicated process of determining the mass of an electron. The results one obtains from an experiment, whether they are instrument readings, counts, “yes” or “no” answers, or values obtained through extensive calculations, are called the **outcomes** of the experiment.

DEFINITION 1. SAMPLE SPACE. *The set of all possible outcomes of an experiment is called the **sample space** and it is usually denoted by the letter S . Each outcome in a sample space is called an **element** of the sample space, or simply a **sample point**.*

If a sample space has a finite number of elements, we may list the elements in the usual set notation; for instance, the sample space for the possible outcomes of one flip of a coin may be written

$$S = \{H, T\}$$

where H and T stand for head and tail. Sample spaces with a large or infinite number of elements are best described by a statement or rule; for example, if the possible outcomes of an experiment are the set of automobiles equipped with satellite radios, the sample space may be written

$$S = \{x | x \text{ is an automobile with a satellite radio}\}$$

This is read “ S is the set of all x such that x is an automobile with a satellite radio.” Similarly, if S is the set of odd positive integers, we write

$$S = \{2k + 1 | k = 0, 1, 2, \dots\}$$

How we formulate the sample space for a given situation will depend on the problem at hand. If an experiment consists of one roll of a die and we are interested in which face is turned up, we would use the sample space

Probability

$$S_1 = \{1, 2, 3, 4, 5, 6\}$$

However, if we are interested only in whether the face turned up is even or odd, we would use the sample space

$$S_2 = \{\text{even, odd}\}$$

This demonstrates that different sample spaces may well be used to describe an experiment. In general, *it is desirable to use sample spaces whose elements cannot be divided (partitioned or separated) into more primitive or more elementary kinds of outcomes*. In other words, *it is preferable that an element of a sample space not represent two or more outcomes that are distinguishable in some way*. Thus, in the preceding illustration S_1 would be preferable to S_2 .

EXAMPLE 2

Describe a sample space that might be appropriate for an experiment in which we roll a pair of dice, one red and one green. (The different colors are used to emphasize that the dice are distinct from one another.)

Solution

The sample space that provides the most information consists of the 36 points given by

$$S_1 = \{(x, y) | x = 1, 2, \dots, 6; y = 1, 2, \dots, 6\}$$

where x represents the number turned up by the red die and y represents the number turned up by the green die. A second sample space, adequate for most purposes (though less desirable in general as it provides less information), is given by

$$S_2 = \{2, 3, 4, \dots, 12\}$$

where the elements are the totals of the numbers turned up by the two dice.

Sample spaces are usually classified according to the number of elements that they contain. In the preceding example the sample spaces S_1 and S_2 contained a **finite** number of elements; but if a coin is flipped until a head appears for the first time, this could happen on the first flip, the second flip, the third flip, the fourth flip, ..., and there are infinitely many possibilities. For this experiment we obtain the sample space

$$S = \{H, TH, TTH, TTTH, TTTTH, \dots\}$$

with an unending sequence of elements. But even here the number of elements can be matched one-to-one with the whole numbers, and in this sense the sample space is said to be **countable**. If a sample space contains a finite number of elements or an infinite though countable number of elements, it is said to be **discrete**.

The outcomes of some experiments are neither finite nor countably infinite. Such is the case, for example, when one conducts an investigation to determine the distance that a certain make of car will travel over a prescribed test course on 5 liters of gasoline. If we assume that distance is a variable that can be measured to any desired degree of accuracy, there is an infinity of possibilities (distances) that cannot be matched one-to-one with the whole numbers. Also, if we want to measure the amount of time it takes for two chemicals to react, the amounts making up the sample space are infinite in number and not countable. Thus, sample spaces need

not be discrete. If a sample space consists of a continuum, such as all the points of a line segment or all the points in a plane, it is said to be **continuous**. Continuous sample spaces arise in practice whenever the outcomes of experiments are measurements of physical properties, such as temperature, speed, pressure, length, . . . , that are measured on continuous scales.

3 Events

In many problems we are interested in results that are not given directly by a specific element of a sample space.

EXAMPLE 3

With reference to the first sample space S_1 on the previous page, describe the event A that the number of points rolled with the die is divisible by 3.

Solution

Among 1, 2, 3, 4, 5, and 6, only 3 and 6 are divisible by 3. Therefore, A is represented by the subset $\{3, 6\}$ of the sample space S_1 .

EXAMPLE 4

With reference to the sample space S_1 of Example 2, describe the event B that the total number of points rolled with the pair of dice is 7.

Solution

Among the 36 possibilities, only (1, 6), (2, 5), (3, 4), (4, 3), (5, 2), and (6, 1) yield a total of 7. So, we write

$$B = \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}$$

Note that in Figure 1 the event of rolling a total of 7 with the two dice is represented by the set of points inside the region bounded by the dotted line.

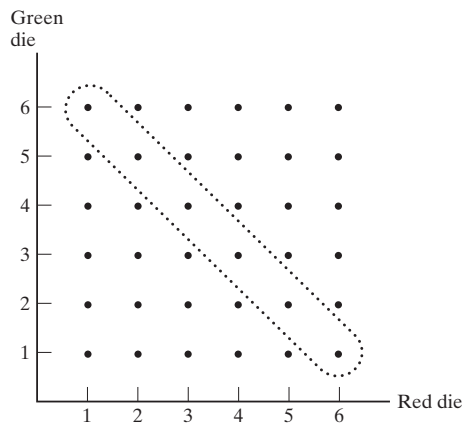


Figure 1. Rolling a total of 7 with a pair of dice.

Probability

In the same way, any event (outcome or result) can be identified with a collection of points, which constitute a subset of an appropriate sample space. Such a subset consists of all the elements of the sample space for which the event occurs, and in probability and statistics we identify the subset with the event.

DEFINITION 2. EVENT. An *event* is a subset of a sample space.

EXAMPLE 5

If someone takes three shots at a target and we care only whether each shot is a hit or a miss, describe a suitable sample space, the elements of the sample space that constitute event M that the person will miss the target three times in a row, and the elements of event N that the person will hit the target once and miss it twice.

Solution

If we let 0 and 1 represent a miss and a hit, respectively, the eight possibilities $(0, 0, 0)$, $(1, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1)$, $(1, 1, 0)$, $(1, 0, 1)$, $(0, 1, 1)$, and $(1, 1, 1)$ may be displayed as in Figure 2. Thus, it can be seen that

$$M = \{(0, 0, 0)\}$$

and

$$N = \{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$$

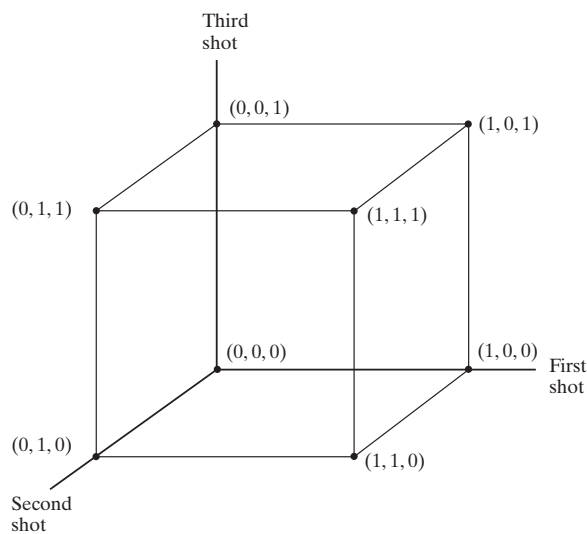


Figure 2. Sample space for Example 5.

EXAMPLE 6

Construct a sample space for the length of the useful life of a certain electronic component and indicate the subset that represents the event F that the component fails before the end of the sixth year.

Solution

If t is the length of the component's useful life in years, the sample space may be written $S = \{t | t \geq 0\}$, and the subset $F = \{t | 0 \leq t < 6\}$ is the event that the component fails before the end of the sixth year.

According to our definition, any event is a subset of an appropriate sample space, but it should be observed that the converse is not necessarily true. For discrete sample spaces, all subsets are events, but in the continuous case some rather abstruse point sets must be excluded for mathematical reasons. This is discussed further in some of the more advanced texts listed among the references at the end of this chapter.

In many problems of probability we are interested in events that are actually combinations of two or more events, formed by taking **unions**, **intersections**, and **complements**. Although the reader must surely be familiar with these terms, let us review briefly that, if A and B are any two subsets of a sample space S , their union $A \cup B$ is the subset of S that contains all the elements that are either in A , in B , or in both; their intersection $A \cap B$ is the subset of S that contains all the elements that are in both A and B ; and the complement A' of A is the subset of S that contains all the elements of S that are not in A . Some of the rules that control the formation of unions, intersections, and complements may be found in Exercises 1 through 4.

Sample spaces and events, particularly relationships among events, are often depicted by means of **Venn diagrams**, in which the sample space is represented by a rectangle, while events are represented by regions within the rectangle, usually by circles or parts of circles. For instance, the shaded regions of the four Venn diagrams of Figure 3 represent, respectively, event A , the complement of event A , the union of events A and B , and the intersection of events A and B . When we are dealing with three events, we usually draw the circles as in Figure 4. Here, the regions are numbered 1 through 8 for easy reference.

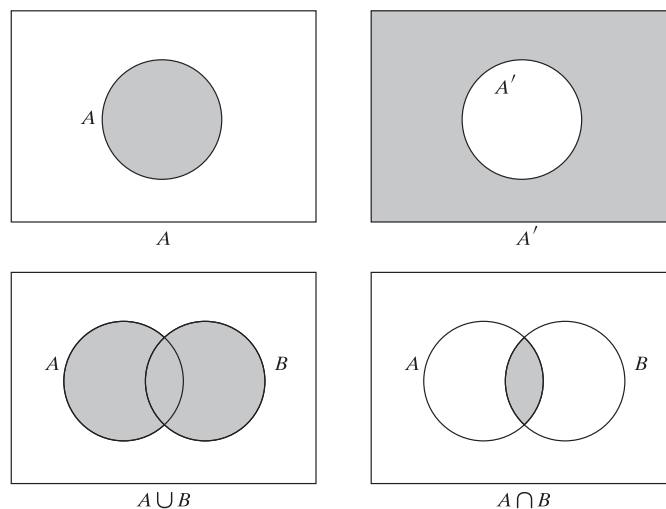


Figure 3. Venn diagrams.

Probability

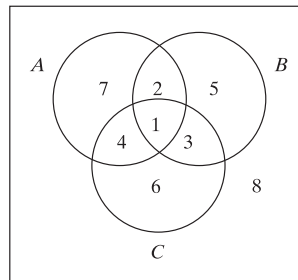


Figure 4. Venn diagram.

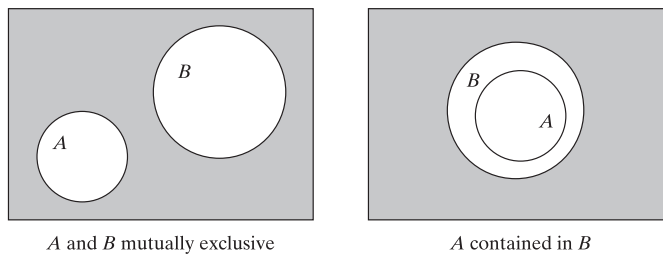


Figure 5. Diagrams showing special relationships among events.

To indicate special relationships among events, we sometimes draw diagrams like those of Figure 5. Here, the one on the left serves to indicate that events A and B are **mutually exclusive**.

DEFINITION 3. MUTUALLY EXCLUSIVE EVENTS. *Two events having no elements in common are said to be **mutually exclusive**.*

When A and B are mutually exclusive, we write $A \cap B = \emptyset$, where \emptyset denotes the **empty set**, which has no elements at all. The diagram on the right serves to indicate that A is contained in B , and symbolically we express this by writing $A \subset B$.

Exercises

1. Use Venn diagrams to verify that
 - (a) $(A \cup B) \cup C$ is the same event as $A \cup (B \cup C)$;
 - (b) $A \cap (B \cup C)$ is the same event as $(A \cap B) \cup (A \cap C)$;
 - (c) $A \cup (B \cap C)$ is the same event as $(A \cup B) \cap (A \cup C)$.
2. Use Venn diagrams to verify the two **De Morgan laws**:
 - (a) $(A \cap B)' = A' \cup B'$;
 - (b) $(A \cup B)' = A' \cap B'$.
3. Use Venn diagrams to verify that
 - (a) $(A \cap B) \cup (A \cap B') = A$;
 - (b) $(A \cap B) \cup (A \cap B') \cup (A' \cap B) = A \cup B$;
 - (c) $A \cup (A' \cap B) = A \cup B$.
4. Use Venn diagrams to verify that if A is contained in B , then $A \cap B = A$ and $A \cap B' = \emptyset$.

4 The Probability of an Event

To formulate the postulates of probability, we shall follow the practice of denoting events by means of capital letters, and we shall write the probability of event A as $P(A)$, the probability of event B as $P(B)$, and so forth. The following postulates of probability apply only to discrete sample spaces, S .

POSTULATE 1 The probability of an event is a nonnegative real number; that is, $P(A) \geq 0$ for any subset A of S .

POSTULATE 2 $P(S) = 1$.

POSTULATE 3 If A_1, A_2, A_3, \dots , is a finite or infinite sequence of mutually exclusive events of S , then

$$P(A_1 \cup A_2 \cup A_3 \cup \dots) = P(A_1) + P(A_2) + P(A_3) + \dots$$

Postulates per se require no proof, but if the resulting theory is to be applied, we must show that the postulates are satisfied when we give probabilities a “real” meaning. Let us illustrate this in connection with the frequency interpretation; the relationship between the postulates and the classical probability concept will be discussed below, while the relationship between the postulates and subjective probabilities is left for the reader to examine in Exercises 16 and 82.

Since proportions are always positive or zero, the first postulate is in complete agreement with the frequency interpretation. The second postulate states indirectly that certainty is identified with a probability of 1; after all, it is always assumed that one of the possibilities in S must occur, and it is to this certain event that we assign a probability of 1. As far as the frequency interpretation is concerned, a probability of 1 implies that the event in question will occur 100 percent of the time or, in other words, that it is certain to occur.

Taking the third postulate in the simplest case, that is, for two mutually exclusive events A_1 and A_2 , it can easily be seen that it is satisfied by the frequency interpretation. If one event occurs, say, 28 percent of the time, another event occurs 39 percent of the time, and the two events cannot both occur at the same time (that is, they are mutually exclusive), then one or the other will occur $28 + 39 = 67$ percent of the time. Thus, the third postulate is satisfied, and the same kind of argument applies when there are more than two mutually exclusive events.

Before we study some of the immediate consequences of the postulates of probability, let us emphasize the point that the three postulates do not tell us how to assign probabilities to events; they merely restrict the ways in which it can be done.

EXAMPLE 7

An experiment has four possible outcomes, A , B , C , and D , that are mutually exclusive. Explain why the following assignments of probabilities are not permissible:

(a) $P(A) = 0.12, P(B) = 0.63, P(C) = 0.45, P(D) = -0.20$;

(b) $P(A) = \frac{9}{120}, P(B) = \frac{45}{120}, P(C) = \frac{27}{120}, P(D) = \frac{46}{120}$.

Solution

(a) $P(D) = -0.20$ violates Postulate 1;

(b) $P(S) = P(A \cup B \cup C \cup D) = \frac{9}{120} + \frac{45}{120} + \frac{27}{120} + \frac{46}{120} = \frac{127}{120} \neq 1$, and this violates Postulate 2.

Probability

Of course, in actual practice probabilities are assigned on the basis of past experience, on the basis of a careful analysis of all underlying conditions, on the basis of subjective judgments, or on the basis of assumptions—sometimes the assumption that all possible outcomes are equiprobable.

To assign a probability measure to a sample space, it is not necessary to specify the probability of each possible subset. This is fortunate, for a sample space with as few as 20 possible outcomes has already $2^{20} = 1,048,576$ subsets, and the number of subsets grows very rapidly when there are 50 possible outcomes, 100 possible outcomes, or more. Instead of listing the probabilities of all possible subsets, we often list the probabilities of the individual outcomes, or sample points of S , and then make use of the following theorem.

THEOREM 1. If A is an event in a discrete sample space S , then $P(A)$ equals the sum of the probabilities of the individual outcomes comprising A .

Proof Let O_1, O_2, O_3, \dots , be the finite or infinite sequence of outcomes that comprise the event A . Thus,

$$A = O_1 \cup O_2 \cup O_3 \cdots$$

and since the individual outcomes, the O 's, are mutually exclusive, the third postulate of probability yields

$$P(A) = P(O_1) + P(O_2) + P(O_3) + \cdots$$

This completes the proof.

To use this theorem, we must be able to assign probabilities to the individual outcomes of experiments. How this is done in some special situations is illustrated by the following examples.

EXAMPLE 8

If we twice flip a balanced coin, what is the probability of getting at least one head?

Solution

The sample space is $S = \{HH, HT, TH, TT\}$, where H and T denote head and tail. Since we assume that the coin is balanced, these outcomes are equally likely and we assign to each sample point the probability $\frac{1}{4}$. Letting A denote the event that we will get at least one head, we get $A = \{HH, HT, TH\}$ and

$$\begin{aligned} P(A) &= P(HH) + P(HT) + P(TH) \\ &= \frac{1}{4} + \frac{1}{4} + \frac{1}{4} \\ &= \frac{3}{4} \end{aligned}$$

EXAMPLE 9

A die is loaded in such a way that each odd number is twice as likely to occur as each even number. Find $P(G)$, where G is the event that a number greater than 3 occurs on a single roll of the die.

Solution

The sample space is $S = \{1, 2, 3, 4, 5, 6\}$. Hence, if we assign probability w to each even number and probability $2w$ to each odd number, we find that $2w + w + 2w + w + 2w + w = 9w = 1$ in accordance with Postulate 2. It follows that $w = \frac{1}{9}$ and

$$P(G) = \frac{1}{9} + \frac{2}{9} + \frac{1}{9} = \frac{4}{9}$$

If a sample space is countably infinite, probabilities will have to be assigned to the individual outcomes by means of a mathematical rule, preferably by means of a formula or equation.

EXAMPLE 10

If, for a given experiment, O_1, O_2, O_3, \dots , is an infinite sequence of outcomes, verify that

$$P(O_i) = \left(\frac{1}{2}\right)^i \quad \text{for } i = 1, 2, 3, \dots$$

is, indeed, a probability measure.

Solution

Since the probabilities are all positive, it remains to be shown that $P(S) = 1$. Getting

$$P(S) = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \dots$$

and making use of the formula for the sum of the terms of an infinite geometric progression, we find that

$$P(S) = \frac{\frac{1}{2}}{1 - \frac{1}{2}} = 1$$

In connection with the preceding example, the word “sum” in Theorem 1 will have to be interpreted so that it includes the value of an infinite series.

The probability measure of Example 10 would be appropriate, for example, if O_i is the event that a person flipping a balanced coin will get a tail for the first time on the i th flip of the coin. Thus, the probability that the first tail will come on the third, fourth, or fifth flip of the coin is

$$\left(\frac{1}{2}\right)^3 + \left(\frac{1}{2}\right)^4 + \left(\frac{1}{2}\right)^5 = \frac{7}{32}$$

and the probability that the first tail will come on an odd-numbered flip of the coin is

$$\left(\frac{1}{2}\right)^1 + \left(\frac{1}{2}\right)^3 + \left(\frac{1}{2}\right)^5 + \dots = \frac{\frac{1}{2}}{1 - \frac{1}{4}} = \frac{2}{3}$$

Probability

Here again we made use of the formula for the sum of the terms of an infinite geometric progression.

If an experiment is such that we can assume equal probabilities for all the sample points, as was the case in Example 8, we can take advantage of the following special case of Theorem 1.

THEOREM 2. If an experiment can result in any one of N different equally likely outcomes, and if n of these outcomes together constitute event A , then the probability of event A is

$$P(A) = \frac{n}{N}$$

Proof Let O_1, O_2, \dots, O_N represent the individual outcomes in S , each with probability $\frac{1}{N}$. If A is the union of n of these mutually exclusive outcomes, and it does not matter which ones, then

$$\begin{aligned} P(A) &= P(O_1 \cup O_2 \cup \dots \cup O_n) \\ &= P(O_1) + P(O_2) + \dots + P(O_n) \\ &= \underbrace{\frac{1}{N} + \frac{1}{N} + \dots + \frac{1}{N}}_{n \text{ terms}} \\ &= \frac{n}{N} \end{aligned}$$

Observe that the formula $P(A) = \frac{n}{N}$ of Theorem 2 is identical with the one for the classical probability concept (see below). Indeed, what we have shown here is that the classical probability concept is consistent with the postulates of probability—it follows from the postulates in the special case where the individual outcomes are all equiprobable.

EXAMPLE 11

A five-card poker hand dealt from a deck of 52 playing cards is said to be a full house if it consists of three of a kind and a pair. If all the five-card hands are equally likely, what is the probability of being dealt a full house?

Solution

The number of ways in which we can be dealt a particular full house, say three kings and two aces, is $\binom{4}{3} \binom{4}{2}$. Since there are 13 ways of selecting the face value for the three of a kind and for each of these there are 12 ways of selecting the face value for the pair, there are altogether

$$n = 13 \cdot 12 \cdot \binom{4}{3} \binom{4}{2}$$

different full houses. Also, the total number of equally likely five-card poker hands is

Probability

$$N = \binom{52}{5}$$

and it follows by Theorem 2 that the probability of getting a full house is

$$P(A) = \frac{n}{N} = \frac{13 \cdot 12 \binom{4}{3} \binom{4}{2}}{\binom{52}{5}} = 0.0014$$

5 Some Rules of Probability

Based on the three postulates of probability, we can derive many other rules that have important applications. Among them, the next four theorems are immediate consequences of the postulates.

THEOREM 3. If A and A' are complementary events in a sample space S , then

$$P(A') = 1 - P(A)$$

Proof In the second and third steps of the proof that follows, we make use of the definition of a complement, according to which A and A' are mutually exclusive and $A \cup A' = S$. Thus, we write

$$\begin{aligned} 1 &= P(S) && \text{(by Postulate 2)} \\ &= P(A \cup A') \\ &= P(A) + P(A') && \text{(by Postulate 3)} \end{aligned}$$

and it follows that $P(A') = 1 - P(A)$.

In connection with the frequency interpretation, this result implies that if an event occurs, say, 37 percent of the time, then it does not occur 63 percent of the time.

THEOREM 4. $P(\emptyset) = 0$ for any sample space S .

Proof Since S and \emptyset are mutually exclusive and $S \cup \emptyset = S$ in accordance with the definition of the empty set \emptyset , it follows that

$$\begin{aligned} P(S) &= P(S \cup \emptyset) \\ &= P(S) + P(\emptyset) && \text{(by Postulate 3)} \end{aligned}$$

and, hence, that $P(\emptyset) = 0$.

It is important to note that it does not necessarily follow from $P(A) = 0$ that $A = \emptyset$. In practice, we often assign 0 probability to events that, in colloquial terms,

would not happen in a million years. For instance, there is the classical example that we assign a probability of 0 to the event that a monkey set loose on a typewriter will type Plato's *Republic* word for word without a mistake. The fact that $P(A) = 0$ does not imply that $A = \emptyset$ is of relevance, especially, in the continuous case.

THEOREM 5. If A and B are events in a sample space S and $A \subset B$, then $P(A) \leq P(B)$.

Proof Since $A \subset B$, we can write

$$B = A \cup (A' \cap B)$$

as can easily be verified by means of a Venn diagram. Then, since A and $A' \cap B$ are mutually exclusive, we get

$$\begin{aligned} P(B) &= P(A) + P(A' \cap B) && \text{(by Postulate 3)} \\ &\geq P(A) && \text{(by Postulate 1)} \end{aligned}$$

In words, this theorem states that if A is a subset of B , then $P(A)$ cannot be greater than $P(B)$. For instance, the probability of drawing a heart from an ordinary deck of 52 playing cards cannot be greater than the probability of drawing a red card. Indeed, the probability of drawing a heart is $\frac{1}{4}$, compared with $\frac{1}{2}$, the probability of drawing a red card.

THEOREM 6. $0 \leq P(A) \leq 1$ for any event A .

Proof Using Theorem 5 and the fact that $\emptyset \subset A \subset S$ for any event A in S , we have

$$P(\emptyset) \leq P(A) \leq P(S)$$

Then, $P(\emptyset) = 0$ and $P(S) = 1$ leads to the result that

$$0 \leq P(A) \leq 1$$

The third postulate of probability is sometimes referred to as the **special addition rule**; it is special in the sense that events A_1, A_2, A_3, \dots , must all be mutually exclusive. For *any* two events A and B , there exists the **general addition rule**, or the **inclusion-exclusion principle**:

THEOREM 7. If A and B are any two events in a sample space S , then

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

Proof Assigning the probabilities a , b , and c to the mutually exclusive events $A \cap B$, $A \cap B'$, and $A' \cap B$ as in the Venn diagram of Figure 6, we find that

$$\begin{aligned} P(A \cup B) &= a + b + c \\ &= (a + b) + (c + a) - a \\ &= P(A) + P(B) - P(A \cap B) \end{aligned}$$

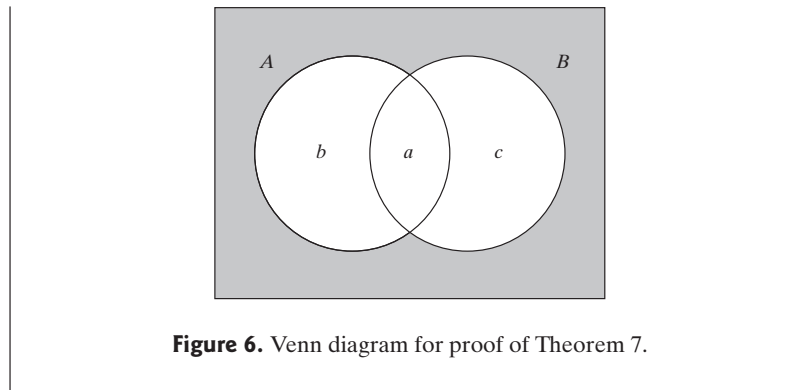


Figure 6. Venn diagram for proof of Theorem 7.

EXAMPLE 12

In a large metropolitan area, the probabilities are 0.86, 0.35, and 0.29, respectively, that a family (randomly chosen for a sample survey) owns a color television set, a HDTV set, or both kinds of sets. What is the probability that a family owns either or both kinds of sets?

Solution

If A is the event that a family in this metropolitan area owns a color television set and B is the event that it owns a HDTV set, we have $P(A) = 0.86$, $P(B) = 0.35$, and $P(A \cap B) = 0.29$; substitution into the formula of Theorem 7 yields

$$\begin{aligned} P(A \cup B) &= 0.86 + 0.35 - 0.29 \\ &= 0.92 \end{aligned}$$

EXAMPLE 13

Near a certain exit of I-17, the probabilities are 0.23 and 0.24, respectively, that a truck stopped at a roadblock will have faulty brakes or badly worn tires. Also, the probability is 0.38 that a truck stopped at the roadblock will have faulty brakes and/or badly worn tires. What is the probability that a truck stopped at this roadblock will have faulty brakes as well as badly worn tires?

Solution

If B is the event that a truck stopped at the roadblock will have faulty brakes and T is the event that it will have badly worn tires, we have $P(B) = 0.23$, $P(T) = 0.24$, and $P(B \cup T) = 0.38$; substitution into the formula of Theorem 7 yields

$$0.38 = 0.23 + 0.24 - P(B \cap T)$$

Solving for $P(B \cap T)$, we thus get

$$P(B \cap T) = 0.23 + 0.24 - 0.38 = 0.09$$

Repeatedly using the formula of Theorem 7, we can generalize this addition rule so that it will apply to any number of events. For instance, for three events we obtain the following theorem.

THEOREM 8. If A , B , and C are any three events in a sample space S , then

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

Proof Writing $A \cup B \cup C$ as $A \cup (B \cup C)$ and using the formula of Theorem 7 twice, once for $P[A \cup (B \cup C)]$ and once for $P(B \cup C)$, we get

$$\begin{aligned} P(A \cup B \cup C) &= P[A \cup (B \cup C)] \\ &= P(A) + P(B \cup C) - P[A \cap (B \cup C)] \\ &= P(A) + P(B) + P(C) - P(B \cap C) \\ &\quad - P[A \cap (B \cup C)] \end{aligned}$$

Then, using the distributive law that the reader was asked to verify in part (b) of Exercise 1, we find that

$$\begin{aligned} P[A \cap (B \cup C)] &= P[(A \cap B) \cup (A \cap C)] \\ &= P(A \cap B) + P(A \cap C) - P[(A \cap B) \cap (A \cap C)] \\ &= P(A \cap B) + P(A \cap C) - P(A \cap B \cap C) \end{aligned}$$

and hence that

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

(In Exercise 12 the reader will be asked to give an alternative proof of this theorem, based on the method used in the text to prove Theorem 7.)

EXAMPLE 14

If a person visits his dentist, suppose that the probability that he will have his teeth cleaned is 0.44, the probability that he will have a cavity filled is 0.24, the probability that he will have a tooth extracted is 0.21, the probability that he will have his teeth cleaned and a cavity filled is 0.08, the probability that he will have his teeth cleaned and a tooth extracted is 0.11, the probability that he will have a cavity filled and a tooth extracted is 0.07, and the probability that he will have his teeth cleaned, a cavity filled, and a tooth extracted is 0.03. What is the probability that a person visiting his dentist will have at least one of these things done to him?

Solution

If C is the event that the person will have his teeth cleaned, F is the event that he will have a cavity filled, and E is the event that he will have a tooth extracted, we are given $P(C) = 0.44$, $P(F) = 0.24$, $P(E) = 0.21$, $P(C \cap F) = 0.08$, $P(C \cap E) = 0.11$, $P(F \cap E) = 0.07$, and $P(C \cap F \cap E) = 0.03$, and substitution into the formula of Theorem 8 yields

$$\begin{aligned} P(C \cup F \cup E) &= 0.44 + 0.24 + 0.21 - 0.08 - 0.11 - 0.07 + 0.03 \\ &= 0.66 \end{aligned}$$

Exercises

5. Use parts (a) and (b) of Exercise 3 to show that

(a) $P(A) \cong P(A \cap B)$;

(b) $P(A) \cong P(A \cup B)$.

6. Referring to Figure 6, verify that

$$P(A \cap B') = P(A) - P(A \cap B)$$

7. Referring to Figure 6 and letting $P(A' \cap B') = d$, verify that

$$P(A' \cap B') = 1 - P(A) - P(B) + P(A \cap B)$$

8. The event that “A or B but not both” will occur can be written as

$$(A \cap B') \cup (A' \cap B)$$

Express the probability of this event in terms of $P(A)$, $P(B)$, and $P(A \cap B)$.

9. Use the formula of Theorem 7 to show that

(a) $P(A \cap B) \leq P(A) + P(B)$;

(b) $P(A \cap B) \geq P(A) + P(B) - 1$.

10. Use the Venn diagram of Figure 7 with the probabilities a, b, c, d, e, f , and g assigned to $A \cap B \cap C$, $A \cap B \cap C'$, ..., and $A \cap B' \cap C'$ to show that if $P(A) = P(B) = P(C) = 1$, then $P(A \cap B \cap C) = 1$. [Hint: Start with the argument that since $P(A) = 1$, it follows that $e = c = f = 0$.]

11. Give an alternative proof of Theorem 7 by making use of the relationships $A \cup B = A \cup (A' \cap B)$ and $B = (A \cap B) \cup (A' \cap B)$.

12. Use the Venn diagram of Figure 7 and the method by which we proved Theorem 7 to prove Theorem 8.

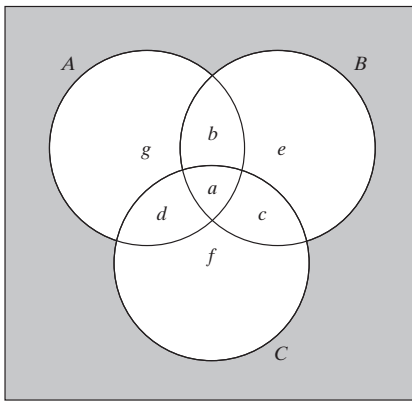


Figure 7. Venn diagram for Exercises 10, 12, and 13.

13. Duplicate the method of proof used in Exercise 12 to show that

$$\begin{aligned} P(A \cup B \cup C \cup D) &= P(A) + P(B) + P(C) + P(D) \\ &\quad - P(A \cap B) - P(A \cap C) - P(A \cap D) \\ &\quad - P(B \cap C) - P(B \cap D) - P(C \cap D) \\ &\quad + P(A \cap B \cap C) + P(A \cap B \cap D) \\ &\quad + P(A \cap C \cap D) + P(B \cap C \cap D) \\ &\quad - P(A \cap B \cap C \cap D) \end{aligned}$$

(Hint: With reference to the Venn diagram of Figure 7, divide each of the eight regions into two parts, designating one to be inside D and the other outside D and letting $a, b, c, d, e, f, g, h, i, j, k, l, m, n, o$, and p be the probabilities associated with the resulting 16 regions.)

14. Prove by induction that

$$P(E_1 \cup E_2 \cup \dots \cup E_n) \leq \sum_{i=1}^n P(E_i)$$

for any finite sequence of events E_1, E_2, \dots , and E_n .

15. The odds that an event will occur are given by the ratio of the probability that the event will occur to the probability that it will not occur, provided neither probability is zero. Odds are usually quoted in terms of positive integers having no common factor. Show that if the odds are A to B that an event will occur, its probability is

$$p = \frac{A}{A + B}$$

16. Subjective probabilities may be determined by exposing persons to risk-taking situations and finding the odds at which they would consider it fair to bet on the outcome. The odds are then converted into probabilities by means of the formula of Exercise 15. For instance, if a person feels that 3 to 2 are fair odds that a business venture will succeed (or that it would be fair to bet \$30 against \$20 that it will succeed), the probability is $\frac{3}{3+2} = 0.6$ that the business venture will succeed. Show that if subjective probabilities are determined in this way, they satisfy

(a) Postulate 1;

(b) Postulate 2.

See also Exercise 82.

6 Conditional Probability

Difficulties can easily arise when probabilities are quoted without specification of the sample space. For instance, if we ask for the probability that a lawyer makes more than \$75,000 per year, we may well get several different answers, and they may all be correct. One of them might apply to all those who are engaged in the private practice of law, another might apply to lawyers employed by corporations, and so forth. Since the choice of the sample space (that is, the set of all possibilities under consideration) is by no means always self-evident, it often helps to use the symbol $P(A|S)$ to denote the **conditional probability** of event A relative to the sample space S or, as we also call it, “the probability of A given S .” The symbol $P(A|S)$ makes it explicit that we are referring to a particular sample space S , and it is preferable to the abbreviated notation $P(A)$ unless the tacit choice of S is clearly understood. It is also preferable when we want to refer to several sample spaces in the same example. If A is the event that a person makes more than \$75,000 per year, G is the event that a person is a law school graduate, L is the event that a person is licensed to practice law, and E is the event that a person is actively engaged in the practice of law, then $P(A|G)$ is the probability that a law school graduate makes more than \$75,000 per year, $P(A|L)$ is the probability that a person licensed to practice law makes more than \$75,000 per year, and $P(A|E)$ is the probability that a person actively engaged in the practice of law makes more than \$75,000 per year.

Some ideas connected with conditional probabilities are illustrated in the following example.

EXAMPLE 15

A consumer research organization has studied the services under warranty provided by the 50 new-car dealers in a certain city, and its findings are summarized in the following table.

	<i>Good service under warranty</i>	<i>Poor service under warranty</i>
<i>In business 10 years or more</i>	16	4
<i>In business less than 10 years</i>	10	20

If a person randomly selects one of these new-car dealers, what is the probability that he gets one who provides good service under warranty? Also, if a person randomly selects one of the dealers who has been in business for 10 years or more, what is the probability that he gets one who provides good service under warranty?

Solution

By “randomly” we mean that, in each case, all possible selections are equally likely, and we can therefore use the formula of Theorem 2. If we let G denote the selection of a dealer who provides good service under warranty, and if we let $n(G)$ denote the number of elements in G and $n(S)$ the number of elements in the whole sample space, we get

$$P(G) = \frac{n(G)}{n(S)} = \frac{16 + 10}{50} = 0.52$$

This answers the first question.

Probability

For the second question, we limit ourselves to the reduced sample space, which consists of the first line of the table, that is, the $16 + 4 = 20$ dealers who have been in business 10 years or more. Of these, 16 provide good service under warranty, and we get

$$P(G|T) = \frac{16}{20} = 0.80$$

where T denotes the selection of a dealer who has been in business 10 years or more. This answers the second question and, as should have been expected, $P(G|T)$ is considerably higher than $P(G)$.

Since the numerator of $P(G|T)$ is $n(T \cap G) = 16$ in the preceding example, the number of dealers who have been in business for 10 years or more and provide good service under warranty, and the denominator is $n(T)$, the number of dealers who have been in business 10 years or more, we can write symbolically

$$P(G|T) = \frac{n(T \cap G)}{n(T)}$$

Then, if we divide the numerator and the denominator by $n(S)$, the total number of new-car dealers in the given city, we get

$$P(G|T) = \frac{\frac{n(T \cap G)}{n(S)}}{\frac{n(T)}{n(S)}} = \frac{P(T \cap G)}{P(T)}$$

and we have, thus, expressed the conditional probability $P(G|T)$ in terms of two probabilities defined for the whole sample space S .

Generalizing from the preceding, let us now make the following definition of conditional probability.

DEFINITION 4. CONDITIONAL PROBABILITY. If A and B are any two events in a sample space S and $P(A) \neq 0$, the **conditional probability** of B given A is

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

EXAMPLE 16

With reference to Example 15, what is the probability that one of the dealers who has been in business less than 10 years will provide good service under warranty?

Solution

Since $P(T' \cap G) = \frac{10}{50} = 0.20$ and $P(T') = \frac{10+20}{50} = 0.60$, substitution into the formula yields

$$P(G|T') = \frac{P(T' \cap G)}{P(T')} = \frac{0.20}{0.60} = \frac{1}{3}$$

Probability

Although we introduced the formula for $P(B|A)$ by means of an example in which the possibilities were all equally likely, this is not a requirement for its use.

EXAMPLE 17

With reference to the loaded die of Example 9, what is the probability that the number of points rolled is a perfect square? Also, what is the probability that it is a perfect square given that it is greater than 3?

Solution

If A is the event that the number of points rolled is greater than 3 and B is the event that it is a perfect square, we have $A = \{4, 5, 6\}$, $B = \{1, 4\}$, and $A \cap B = \{4\}$. Since the probabilities of rolling a 1, 2, 3, 4, 5, or 6 with the die are $\frac{2}{9}$, $\frac{1}{9}$, $\frac{2}{9}$, $\frac{1}{9}$, $\frac{2}{9}$, and $\frac{1}{9}$, we find that the answer to the first question is

$$P(B) = \frac{2}{9} + \frac{1}{9} = \frac{1}{3}$$

To determine $P(B|A)$, we first calculate

$$P(A \cap B) = \frac{1}{9} \quad \text{and} \quad P(A) = \frac{1}{9} + \frac{2}{9} + \frac{1}{9} = \frac{4}{9}$$

Then, substituting into the formula of Definition 4, we get

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{\frac{1}{9}}{\frac{4}{9}} = \frac{1}{4}$$

To verify that the formula of Definition 4 has yielded the “right” answer in the preceding example, we have only to assign probability v to the two even numbers in the reduced sample space A and probability $2v$ to the odd number, such that the sum of the three probabilities is equal to 1. We thus have $v + 2v + v = 1$, $v = \frac{1}{4}$, and, hence, $P(B|A) = \frac{1}{4}$ as before.

EXAMPLE 18

A manufacturer of airplane parts knows from past experience that the probability is 0.80 that an order will be ready for shipment on time, and it is 0.72 that an order will be ready for shipment on time and will also be delivered on time. What is the probability that such an order will be delivered on time given that it was ready for shipment on time?

Solution

If we let R stand for the event that an order is ready for shipment on time and D be the event that it is delivered on time, we have $P(R) = 0.80$ and $P(R \cap D) = 0.72$, and it follows that

$$P(D|R) = \frac{P(R \cap D)}{P(R)} = \frac{0.72}{0.80} = 0.90$$

Probability

Thus, 90 percent of the shipments will be delivered on time provided they are shipped on time. Note that $P(R|D)$, the probability that a shipment that is delivered on time was also ready for shipment on time, cannot be determined without further information; for this purpose we would also have to know $P(D)$.

If we multiply the expressions on both sides of the formula of Definition 4 by $P(A)$, we obtain the following **multiplication rule**.

THEOREM 9. If A and B are any two events in a sample space S and $P(A) \neq 0$, then

$$P(A \cap B) = P(A) \cdot P(B|A)$$

In words, the probability that A and B will both occur is the product of the probability of A and the conditional probability of B given A . Alternatively, if $P(B) \neq 0$, the probability that A and B will both occur is the product of the probability of B and the conditional probability of A given B ; symbolically,

$$P(A \cap B) = P(B) \cdot P(A|B)$$

To derive this alternative multiplication rule, we interchange A and B in the formula of Theorem 9 and make use of the fact that $A \cap B = B \cap A$.

EXAMPLE 19

If we randomly pick two television sets in succession from a shipment of 240 television sets of which 15 are defective, what is the probability that they will both be defective?

Solution

If we assume equal probabilities for each selection (which is what we mean by “randomly” picking the sets), the probability that the first set will be defective is $\frac{15}{240}$, and the probability that the second set will be defective given that the first set is defective is $\frac{14}{239}$. Thus, the probability that both sets will be defective is $\frac{15}{240} \cdot \frac{14}{239} = \frac{7}{1,912}$. This assumes that we are **sampling without replacement**; that is, the first set is not replaced before the second set is selected.

EXAMPLE 20

Find the probabilities of randomly drawing two aces in succession from an ordinary deck of 52 playing cards if we sample

- (a) without replacement;
- (b) with replacement.

Solution

- (a) If the first card is not replaced before the second card is drawn, the probability of getting two aces in succession is

$$\frac{4}{52} \cdot \frac{3}{51} = \frac{1}{221}$$

Probability

- (b) If the first card is replaced before the second card is drawn, the corresponding probability is

$$\frac{4}{52} \cdot \frac{4}{52} = \frac{1}{169}$$

In the situations described in the two preceding examples there is a definite temporal order between the two events A and B . In general, this need not be the case when we write $P(A|B)$ or $P(B|A)$. For instance, we could ask for the probability that the first card drawn was an ace given that the second card drawn (without replacement) is an ace—the answer would also be $\frac{3}{51}$.

Theorem 9 can easily be generalized so that it applies to more than two events; for instance, for three events we have the following theorem.

THEOREM 10. If A , B , and C are any three events in a sample space S such that $P(A \cap B) \neq 0$, then

$$P(A \cap B \cap C) = P(A) \cdot P(B|A) \cdot P(C|A \cap B)$$

Proof Writing $A \cap B \cap C$ as $(A \cap B) \cap C$ and using the formula of Theorem 9 twice, we get

$$\begin{aligned} P(A \cap B \cap C) &= P[(A \cap B) \cap C] \\ &= P(A \cap B) \cdot P(C|A \cap B) \\ &= P(A) \cdot P(B|A) \cdot P(C|A \cap B) \end{aligned}$$

EXAMPLE 21

A box of fuses contains 20 fuses, of which 5 are defective. If 3 of the fuses are selected at random and removed from the box in succession without replacement, what is the probability that all 3 fuses are defective?

Solution

If A is the event that the first fuse is defective, B is the event that the second fuse is defective, and C is the event that the third fuse is defective, then $P(A) = \frac{5}{20}$, $P(B|A) = \frac{4}{19}$, $P(C|A \cap B) = \frac{3}{18}$, and substitution into the formula yields

$$\begin{aligned} P(A \cap B \cap C) &= \frac{5}{20} \cdot \frac{4}{19} \cdot \frac{3}{18} \\ &= \frac{1}{114} \end{aligned}$$

Further generalization of Theorems 9 and 10 to k events is straightforward, and the resulting formula can be proved by mathematical induction.

7 Independent Events

Informally speaking, two events A and B are **independent** if the occurrence or nonoccurrence of either one does not affect the probability of the occurrence of the other. For instance, in the preceding example the selections would all have been independent had each fuse been replaced before the next one was selected; the probability of getting a defective fuse would have remained $\frac{5}{20}$.

Symbolically, two events A and B are independent if $P(B|A) = P(B)$ and $P(A|B) = P(A)$, and it can be shown that either of these equalities implies the other when both of the conditional probabilities exist, that is, when neither $P(A)$ nor $P(B)$ equals zero (see Exercise 21).

Now, if we substitute $P(B)$ for $P(B|A)$ into the formula of Theorem 9, we get

$$\begin{aligned} P(A \cap B) &= P(A) \cdot P(B|A) \\ &= P(A) \cdot P(B) \end{aligned}$$

and we shall use this as our formal definition of independence.

DEFINITION 5. INDEPENDENCE. *Two events A and B are **independent** if and only if*

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

Reversing the steps, we can also show that Definition 5 implies the definition of independence that we gave earlier.

If two events are not independent, they are said to be **dependent**. In the derivation of the formula of Definition 5, we assume that $P(B|A)$ exists and, hence, that $P(A) \neq 0$. For mathematical convenience, we shall let the definition apply also when $P(A) = 0$ and/or $P(B) = 0$.

EXAMPLE 22

A coin is tossed three times and the eight possible outcomes, HHH, HHT, HTH, THH, HTT, THT, TTH, and TTT, are assumed to be equally likely. If A is the event that a head occurs on each of the first two tosses, B is the event that a tail occurs on the third toss, and C is the event that exactly two tails occur in the three tosses, show that

- (a) events A and B are independent;
- (b) events B and C are dependent.

Solution

Since

$$\begin{aligned} A &= \{HHH, HHT\} \\ B &= \{HHT, HTT, THT, TTT\} \\ C &= \{HTT, THT, TTH\} \\ A \cap B &= \{HHT\} \\ B \cap C &= \{HTT, THT\} \end{aligned}$$

Probability

the assumption that the eight possible outcomes are all equiprobable yields $P(A) = \frac{1}{4}$, $P(B) = \frac{1}{2}$, $P(C) = \frac{3}{8}$, $P(A \cap B) = \frac{1}{8}$, and $P(B \cap C) = \frac{1}{4}$.

- (a) Since $P(A) \cdot P(B) = \frac{1}{4} \cdot \frac{1}{2} = \frac{1}{8} = P(A \cap B)$, events A and B are independent.
- (b) Since $P(B) \cdot P(C) = \frac{1}{2} \cdot \frac{3}{8} = \frac{3}{16} \neq P(B \cap C)$, events B and C are not independent.

In connection with Definition 5, it can be shown that if A and B are independent, then so are A and B' , A' and B , and A' and B' . For instance, consider the following theorem.

THEOREM 11. If A and B are independent, then A and B' are also independent.

Proof Since $A = (A \cap B) \cup (A \cap B')$, as the reader was asked to show in part (a) of Exercise 3, $A \cap B$ and $A \cap B'$ are mutually exclusive, and A and B are independent by assumption, we have

$$\begin{aligned} P(A) &= P[(A \cap B) \cup (A \cap B')] \\ &= P(A \cap B) + P(A \cap B') \\ &= P(A) \cdot P(B) + P(A \cap B') \end{aligned}$$

It follows that

$$\begin{aligned} P(A \cap B') &= P(A) - P(A) \cdot P(B) \\ &= P(A) \cdot [1 - P(B)] \\ &= P(A) \cdot P(B') \end{aligned}$$

and hence that A and B' are independent.

In Exercises 22 and 23 the reader will be asked to show that if A and B are independent, then A' and B are independent and so are A' and B' , and if A and B are dependent, then A and B' are dependent.

To extend the concept of independence to more than two events, let us make the following definition.

DEFINITION 6. INDEPENDENCE OF MORE THAN TWO EVENTS. Events A_1, A_2, \dots , and A_k are **independent** if and only if the probability of the intersections of any 2, 3, \dots , or k of these events equals the product of their respective probabilities.

For three events A , B , and C , for example, independence requires that

$$\begin{aligned} P(A \cap B) &= P(A) \cdot P(B) \\ P(A \cap C) &= P(A) \cdot P(C) \\ P(B \cap C) &= P(B) \cdot P(C) \end{aligned}$$

and

$$P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)$$

It is of interest to note that three or more events can be **pairwise independent** without being independent.

EXAMPLE 23

Figure 8 shows a Venn diagram with probabilities assigned to its various regions. Verify that A and B are independent, A and C are independent, and B and C are independent, but A , B , and C are not independent.

Solution

As can be seen from the diagram, $P(A) = P(B) = P(C) = \frac{1}{2}$, $P(A \cap B) = P(A \cap C) = P(B \cap C) = \frac{1}{4}$, and $P(A \cap B \cap C) = \frac{1}{4}$. Thus,

$$P(A) \cdot P(B) = \frac{1}{4} = P(A \cap B)$$

$$P(A) \cdot P(C) = \frac{1}{4} = P(A \cap C)$$

$$P(B) \cdot P(C) = \frac{1}{4} = P(B \cap C)$$

but

$$P(A) \cdot P(B) \cdot P(C) = \frac{1}{8} \neq P(A \cap B \cap C)$$

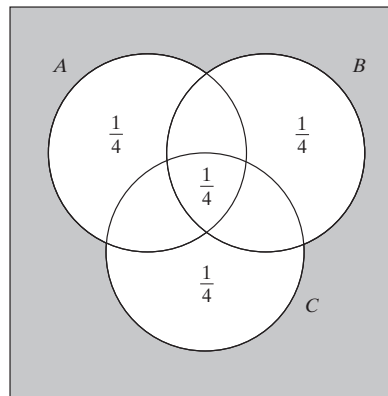


Figure 8. Venn diagram for Example 23.

Incidentally, the preceding example can be given a “real” interpretation by considering a large room that has three separate switches controlling the ceiling lights. These lights will be on when all three switches are “up” and hence also when one of the switches is “up” and the other two are “down.” If A is the event that the first switch is “up,” B is the event that the second switch is “up,” and C is the event that the third switch is “up,” the Venn diagram of Figure 8 shows a possible set of probabilities associated with the switches being “up” or “down” when the ceiling lights are on.

It can also happen that $P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)$ without A , B , and C being pairwise independent—this the reader will be asked to verify in Exercise 24.

Of course, if we are given that certain events are independent, the probability that they will all occur is simply the product of their respective probabilities.

EXAMPLE 24

Find the probabilities of getting

- (a) three heads in three random tosses of a balanced coin;
- (b) four sixes and then another number in five random rolls of a balanced die.

Solution

- (a) The probability of a head on each toss is $\frac{1}{2}$ and the three outcomes are independent. Thus we can multiply, obtaining

$$\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$$

- (b) The probability of a six on each toss is $\frac{1}{6}$; thus the probability of tossing a number other than 6 is $\frac{5}{6}$. Inasmuch as the tosses are independent, we can multiply the respective probabilities to obtain

$$\frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{5}{6} = \frac{5}{7,776}$$

8 Bayes' Theorem

In many situations the outcome of an experiment depends on what happens in various intermediate stages. The following is a simple example in which there is one intermediate stage consisting of two alternatives:

EXAMPLE 25

The completion of a construction job may be delayed because of a strike. The probabilities are 0.60 that there will be a strike, 0.85 that the construction job will be completed on time if there is no strike, and 0.35 that the construction job will be completed on time if there is a strike. What is the probability that the construction job will be completed on time?

Solution

If A is the event that the construction job will be completed on time and B is the event that there will be a strike, we are given $P(B) = 0.60$, $P(A|B') = 0.85$, and $P(A|B) = 0.35$. Making use of the formula of part (a) of Exercise 3, the fact that $A \cap B$ and $A \cap B'$ are mutually exclusive, and the alternative form of the multiplication rule, we can write

$$\begin{aligned} P(A) &= P[(A \cap B) \cup (A \cap B')] \\ &= P(A \cap B) + P(A \cap B') \\ &= P(B) \cdot P(A|B) + P(B') \cdot P(A|B') \end{aligned}$$

Probability

Then, substituting the given numerical values, we get

$$\begin{aligned} P(A) &= (0.60)(0.35) + (1 - 0.60)(0.85) \\ &= 0.55 \end{aligned}$$

An immediate generalization of this kind of situation is the case where the intermediate stage permits k different alternatives (whose occurrence is denoted by B_1, B_2, \dots, B_k). It requires the following theorem, sometimes called the **rule of total probability** or the **rule of elimination**.

THEOREM 12. If the events B_1, B_2, \dots , and B_k constitute a partition of the sample space S and $P(B_i) \neq 0$ for $i = 1, 2, \dots, k$, then for any event A in S

$$P(A) = \sum_{i=1}^k P(B_i) \cdot P(A|B_i)$$

The B 's constitute a partition of the sample space if they are pairwise mutually exclusive and if their union equals S . A formal proof of Theorem 12 consists, essentially, of the same steps we used in Example 25, and it is left to the reader in Exercise 32.

EXAMPLE 26

The members of a consulting firm rent cars from three rental agencies: 60 percent from agency 1, 30 percent from agency 2, and 10 percent from agency 3. If 9 percent of the cars from agency 1 need an oil change, 20 percent of the cars from agency 2 need an oil change, and 6 percent of the cars from agency 3 need an oil change, what is the probability that a rental car delivered to the firm will need an oil change?

Solution

If A is the event that the car needs an oil change, and B_1, B_2 , and B_3 are the events that the car comes from rental agencies 1, 2, or 3, we have $P(B_1) = 0.60$, $P(B_2) = 0.30$, $P(B_3) = 0.10$, $P(A|B_1) = 0.09$, $P(A|B_2) = 0.20$, and $P(A|B_3) = 0.06$. Substituting these values into the formula of Theorem 12, we get

$$\begin{aligned} P(A) &= (0.60)(0.09) + (0.30)(0.20) + (0.10)(0.06) \\ &= 0.12 \end{aligned}$$

Thus, 12 percent of all the rental cars delivered to this firm will need an oil change.

With reference to the preceding example, suppose that we are interested in the following question: If a rental car delivered to the consulting firm needs an oil change, what is the probability that it came from rental agency 2? To answer questions of this kind, we need the following theorem, called **Bayes' theorem**:

THEOREM 13. If B_1, B_2, \dots, B_k constitute a partition of the sample space S and $P(B_i) \neq 0$ for $i = 1, 2, \dots, k$, then for any event A in S such that $P(A) \neq 0$

$$P(B_r|A) = \frac{P(B_r) \cdot P(A|B_r)}{\sum_{i=1}^k P(B_i) \cdot P(A|B_i)}$$

for $r = 1, 2, \dots, k$.

In words, the probability that event A was reached via the r th branch of the tree diagram of Figure 9, given that it was reached via one of its k branches, is the *ratio* of the probability associated with the r th branch to the sum of the probabilities associated with all k branches of the tree.

Proof Writing $P(B_r|A) = \frac{P(A \cap B_r)}{P(A)}$ in accordance with the definition of conditional probability, we have only to substitute $P(B_r) \cdot P(A|B_r)$ for $P(A \cap B_r)$ and the formula of Theorem 12 for $P(A)$.

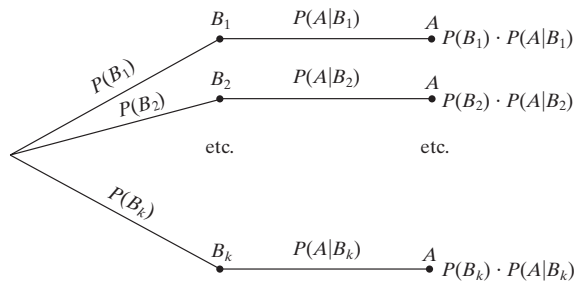


Figure 9. Tree diagram for Bayes' theorem.

EXAMPLE 27

With reference to Example 26, if a rental car delivered to the consulting firm needs an oil change, what is the probability that it came from rental agency 2?

Solution

Substituting the probabilities on the previous page into the formula of Theorem 13, we get

$$\begin{aligned} P(B_2|A) &= \frac{(0.30)(0.20)}{(0.60)(0.09) + (0.30)(0.20) + (0.10)(0.06)} \\ &= \frac{0.060}{0.120} \\ &= 0.5 \end{aligned}$$

Observe that although only 30 percent of the cars delivered to the firm come from agency 2, 50 percent of those requiring an oil change come from that agency.

EXAMPLE 28

A rare but serious disease, D , has been found in 0.01 percent of a certain population. A test has been developed that will be positive, p , for 98 percent of those who have the disease and be positive for only 3 percent of those who do not have the disease. Find the probability that a person tested as positive does not have the disease.

Solution

Let \bar{D} and \bar{p} represent the events that a person randomly selected from the given population, respectively, does not have the disease and is found negative for the disease by the test. Substituting the given probabilities into the formula of Theorem 13, we get

$$P(\bar{D}|p) = \frac{P(\bar{D})P(p|\bar{D})}{P(D)P(p|D) + P(\bar{D})P(p|\bar{D})} = \frac{0.9999 \cdot 0.03}{0.0001 \cdot 0.98 + 0.9999 \cdot 0.03} = 0.997$$

This example demonstrates the near impossibility of finding a test for a rare disease that does not have an unacceptably high probability of false positives.

Although Bayes' theorem follows from the postulates of probability and the definition of conditional probability, it has been the subject of extensive controversy. There can be no question about the validity of Bayes' theorem, but considerable arguments have been raised about the assignment of the **prior probabilities** $P(B_i)$. Also, a good deal of mysticism surrounds Bayes' theorem because it entails a "backward," or "inverse," sort of reasoning, that is, reasoning "from effect to cause." For instance, in Example 27, needing an oil change is the effect and coming from agency 2 of is a possible cause.

Exercises

17. Show that the postulates of probability are satisfied by conditional probabilities. In other words, show that if $P(B) \neq 0$, then

(a) $P(A|B) \geq 0$;

(b) $P(B|B) = 1$;

(c) $P(A_1 \cup A_2 \cup \dots | B) = P(A_1|B) + P(A_2|B) + \dots$ for any sequence of mutually exclusive events A_1, A_2, \dots

18. Show by means of numerical examples that $P(B|A) + P(B|A')$

(a) may be equal to 1;

(b) need not be equal to 1.

19. Duplicating the method of proof of Theorem 10, show that $P(A \cap B \cap C \cap D) = P(A) \cdot P(B|A) \cdot P(C|A \cap B) \cdot P(D|A \cap B \cap C)$ provided that $P(A \cap B \cap C) \neq 0$.

20. Given three events A , B , and C such that $P(A \cap B \cap C) \neq 0$ and $P(C|A \cap B) = P(C|B)$, show that $P(A|B \cap C) = P(A|B)$.

21. Show that if $P(B|A) = P(B)$ and $P(B) \neq 0$, then $P(A|B) = P(A)$.

22. Show that if events A and B are independent, then

(a) events A' and B are independent;

(b) events A' and B' are independent.

23. Show that if events A and B are dependent, then events A and B' are dependent.

24. Refer to Figure 10 to show that $P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)$ does not necessarily imply that A , B , and C are all pairwise independent.

25. Refer to Figure 10 to show that if A is independent of B and A is independent of C , then B is not necessarily independent of C .

26. Refer to Figure 10 to show that if A is independent of B and A is independent of C , then A is not necessarily independent of $B \cup C$.

27. If events A , B , and C are independent, show that

(a) A and $B \cap C$ are independent;

(b) A and $B \cup C$ are independent.

28. If $P(A|B) < P(A)$, prove that $P(B|A) < P(B)$.

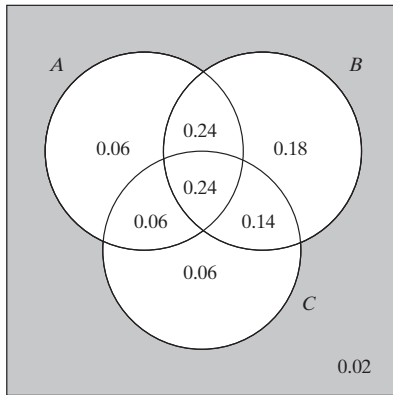


Figure 10. Diagram for Exercises 24, 25, and 26.

29. If A_1, A_2, \dots, A_n are independent events, prove that
- $$P(A_1 \cup A_2 \cup \dots \cup A_n) = 1 - \{1 - P(A_1)\} \cdot \{1 - P(A_2)\} \dots \{1 - P(A_n)\}$$

30. Show that $2^k - k - 1$ conditions must be satisfied for k events to be independent.

31. For any event A , show that A and \emptyset are independent.

32. Prove Theorem 12 by making use of the following generalization of the distributive law given in part (b) of Exercise 1:

$$A \cap (B_1 \cup B_2 \cup \dots \cup B_k) = (A \cap B_1) \cup (A \cap B_2) \cup \dots \cup (A \cap B_k)$$

33. Suppose that a die has n sides numbered $i = 1, 2, \dots, n$. Assume that the probability of it coming up on the side numbered i is the same for each value of i . The die is rolled n times (assume independence) and a “match” is defined to be the occurrence of side i on the i th roll. Prove that the probability of at least one match is given by

$$1 - \left(\frac{n-1}{n}\right)^n = 1 - \left(1 - \frac{1}{n}\right)^n$$

34. Show that $P(A \cup B) \geq 1 - P(A') - P(B')$ for any two events A and B defined in the sample space S . (Hint: Use Venn diagrams.)

9 The Theory in Practice

The word “probability” is a part of everyday language, but it is difficult to define this word without using the word “probable” or its synonym “likely” in the definition.* To illustrate, Webster’s *Third New International Dictionary* defines “probability” as “the quality or state of being probable.” If the concept of probability is to be used in mathematics and scientific applications, we require a more exact, less circular, definition.

The postulates of probability given in Section 4 satisfy this criterion. Together with the rules given in Section 5, this definition lends itself to calculations of probabilities that “make sense” and that can be verified experimentally. The entire theory of statistics is based on the notion of probability. It seems remarkable that the entire structure of probability and, therefore of statistics, can be built on the relatively straightforward foundation given in this chapter.

Probabilities were first considered in games of chance, or gambling. Players of various games of chance observed that there seemed to be “rules” that governed the roll of dice or the results of spinning a roulette wheel. Some of them went as far as to postulate some of these rules entirely on the basis of experience. But differences arose among gamblers about probabilities, and they brought their questions to the noted mathematicians of their day. With this motivation, the modern theory of probability began to be developed.

Motivated by problems associated with games of chance, the theory of probability first was developed under the assumption of **equal likelihood**, expressed in Theorem 2. Under this assumption one only had to count the number of “successful” outcomes and divide by the total number of “possible” outcomes to arrive at the probability of an event.

The assumption of equal likelihood fails when we attempt, for example, to find the probability that a trifecta at the race track will pay off. Here, the different horses have different probabilities of winning, and we are forced to rely on a different method of evaluating probabilities. It is common to take into account the various

*From MERRIAM-WEBSTER’S COLLEGIATE DICTIONARY, ELEVENTH EDITION. Copyright © 2012 by Merriam-Webster, Incorporated (www.Merriam-Webster.com). Reprinted with permission.

horses' records in previous races, calculating each horse's probability of winning by dividing its number of wins by the number of starts. This idea gives rise to the **frequency interpretation** of probabilities, which interprets the probability of an event to be the proportion of times the event has occurred in a long series of repeated experiments. Application of the frequency interpretation requires a well-documented history of the outcomes of an event over a large number of experimental trials. In the absence of such a history, a series of experiments can be planned and their results observed. For example, the probability that a lot of manufactured items will have at most three defectives is estimated to be 0.90 if, in 90 percent of many previous lots *produced to the same specifications by the same process*, the number of defectives was three or less.

A more recently employed method of calculating probabilities is called the **subjective method**. Here, a personal or subjective assessment is made of the probability of an event which is difficult or impossible to estimate in any other way. For example, the probability that the major stock market indexes will go up in a given future period of time cannot be estimated very well using the frequency interpretation because economic and world conditions rarely replicate themselves very closely. As another example, juries use this method when determining the guilt or innocence of a defendant "beyond a reasonable doubt." Subjective probabilities should be used only when all other methods fail, and then only with a high level of skepticism.

An important application of probability theory relates to the theory of **reliability**. The reliability of a component or system can be defined as follows.

DEFINITION 7. RELIABILITY. *The reliability of a product is the probability that it will function within specified limits for a specified period of time under specified environmental conditions.*

Thus, the reliability of a "standard equipment" automobile tire is close to 1 for 10,000 miles of operation on a passenger car traveling within the speed limits on paved roads, but it is close to zero for even short distances at the Indianapolis "500."

The reliability of a system of components can be calculated from the reliabilities of the individual components if the system consists entirely of components connected in series, or in parallel, or both. A **series system** is one in which all components are so interrelated that the entire system will fail if any one (or more) of its components fails. A **parallel system** will fail only if all its components fail. An example of a series system is a string of Christmas lights connected electrically "in series." If one bulb fails, the entire string will fail to light. Parallel systems are sometimes called "redundant" systems. For example, if the hydraulic system on a commercial aircraft that lowers the landing wheels fails, they may be lowered manually with a crank.

We shall assume that the components connected in a series system are independent; that is, the performance of one part does not affect the reliability of the others. Under this assumption, the reliability of a parallel system is given by an extension of Definition 5. Thus, we have the following theorem.

THEOREM 14. The **reliability of a series system** consisting of n independent components is given by

$$R_s = \prod_{i=1}^n R_i$$

where R_i is the reliability of the i th component.

Proof The proof follows immediately by iterating in Definition 5.

Probability

Theorem 14 vividly demonstrates the effect of increased complexity on reliability. For example, if a series system has 5 components, each with a reliability of 0.970, the reliability of the entire system is only $(0.970)^5 = 0.859$. If the system complexity were increased so it now has 10 such components, the reliability would be reduced to $(0.970)^{10} = 0.738$.

One way to improve the reliability of a series system is to introduce parallel redundancy by replacing some or all of its components by several components connected in parallel. If a system consists of n independent components connected in parallel, it will fail to function only if all components fail. Thus, for the i th component, the probability of failure is $F_i = 1 - R_i$, called the “unreliability” of the component. Again applying Definition 5, we obtain the following theorem.

THEOREM 15. The **reliability of a parallel system** consisting of n independent components is given by

$$R_p = 1 - \prod_{i=1}^n (1 - R_i)$$

Proof The proof of this theorem is identical to that of Theorem 14, with $(1 - R_i)$ replacing R_i .

EXAMPLE 29

Consider the system diagramed in Figure 11, which consists of eight components having the reliabilities shown in the figure. Find the reliability of the system.

Solution

The parallel subsystem C, D, E can be replaced by an equivalent component, C' having the reliability $1 - (1 - 0.70)^3 = 0.973$. Likewise, F, G can be replaced by F' having the reliability $1 - (1 - 0.75)^2 = 0.9375$. Thus, the system is reduced to the parallel system A, B, C', F', H , having the reliability $(0.95)(0.99)(0.973)(0.9375)(0.90) = 0.772$.

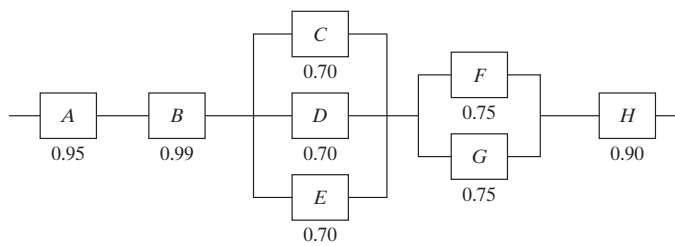


Figure 11. Combination of series and parallel systems.

Applied Exercises

SECS. 1–3

35. If $S = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$, $A = \{1, 3, 5, 7\}$, $B = \{6, 7, 8, 9\}$, $C = \{2, 4, 8\}$, and $D = \{1, 5, 9\}$, list the elements of the subsets of S corresponding to the following events:

- (a) $A' \cap B$; (b) $(A' \cap B) \cap C$; (c) $B' \cup C$;
 (d) $(B' \cup C) \cap D$; (e) $A' \cap C$; (f) $(A' \cap C) \cap D$.

36. An electronics firm plans to build a research laboratory in Southern California, and its management has to decide between sites in Los Angeles, San Diego, Long Beach, Pasadena, Santa Barbara, Anaheim, Santa Monica, and Westwood. If A represents the event that they will choose a site in San Diego or Santa Barbara, B represents the event that they will choose a site in San Diego or Long Beach, C represents the event that they will choose a site in Santa Barbara or Anaheim, and D represents the event that they will choose a site in Los Angeles or Santa Barbara, list the elements of each of the following subsets of the sample space, which consists of the eight site selections:

- (a) A' ; (b) D' ; (c) $C \cap D$;
 (d) $B \cap C$; (e) $B \cup C$; (f) $A \cup B$;
 (g) $C \cup D$; (h) $(B \cup C)'$; (i) $B' \cap C'$.

37. Among the eight cars that a dealer has in his showroom, Car 1 is new and has air-conditioning, power steering, and bucket seats; Car 2 is one year old and has air-conditioning, but neither power steering nor bucket seats; Car 3 is two years old and has air-conditioning and power steering, but no bucket seats; Car 4 is three years old and has air-conditioning, but neither power steering nor bucket seats; Car 5 is new and has no air-conditioning, no power steering, and no bucket seats; Car 6 is one year old and has power steering, but neither air-conditioning nor bucket seats; Car 7 is two years old and has no air-conditioning, no power steering, and no bucket seats; and Car 8 is three years old, and has no air-conditioning, but has power steering as well as bucket seats. If a customer buys one of these cars and the event that he chooses a new car, for example, is represented by the set {Car 1, Car 5}, indicate similarly the sets that represent the events that

- (a) he chooses a car without air-conditioning;
 (b) he chooses a car without power steering;
 (c) he chooses a car with bucket seats;
 (d) he chooses a car that is either two or three years old.

38. With reference to Exercise 37, state in words what kind of car the customer will choose, if his choice is given by

- (a) the complement of the set of part (a);
 (b) the union of the sets of parts (b) and (c);

- (c) the intersection of the sets of parts (c) and (d);
 (d) the intersection of parts (b) and (c) of this exercise.

39. If Ms. Brown buys one of the houses advertised for sale in a Seattle newspaper (on a given Sunday), T is the event that the house has three or more baths, U is the event that it has a fireplace, V is the event that it costs more than \$200,000, and W is the event that it is new, describe (in words) each of the following events:

- (a) T' ; (b) U' ; (c) V' ;
 (d) W' ; (e) $T \cap U$; (f) $T \cap V$;
 (g) $U' \cap V$; (h) $V \cup W$; (i) $V' \cup W$;
 (j) $T \cup U$; (k) $T \cup V$; (l) $V \cap W$.

40. A resort hotel has two station wagons, which it uses to shuttle its guests to and from the airport. If the larger of the two station wagons can carry five passengers and the smaller can carry four passengers, the point $(0, 3)$ represents the event that at a given moment the larger station wagon is empty while the smaller one has three passengers, the point $(4, 2)$ represents the event that at the given moment the larger station wagon has four passengers while the smaller one has two passengers, ..., draw a figure showing the 30 points of the corresponding sample space. Also, if E stands for the event that at least one of the station wagons is empty, F stands for the event that together they carry two, four, or six passengers, and G stands for the event that each carries the same number of passengers, list the points of the sample space that correspond to each of the following events:

- (a) E ; (b) F ; (c) G ;
 (d) $E \cup F$; (e) $E \cap F$; (f) $F \cup G$;
 (g) $E \cup F'$; (h) $E \cap G'$; (i) $F' \cap E'$.

41. A coin is tossed once. Then, if it comes up heads, a die is thrown once; if the coin comes up tails, it is tossed twice more. Using the notation in which $(H, 2)$, for example, denotes the event that the coin comes up heads and then the die comes up 2, and (T, T, T) denotes the event that the coin comes up tails three times in a row, list

- (a) the 10 elements of the sample space S ;
 (b) the elements of S corresponding to event A that exactly one head occurs;
 (c) the elements of S corresponding to event B that at least two tails occur or a number greater than 4 occurs.

42. An electronic game contains three components arranged in the series-parallel circuit shown in Figure 12. At any given time, each component may or may not be operative, and the game will operate only if there is a continuous circuit from P to Q . Let A be the event that the game will operate; let B be the event that

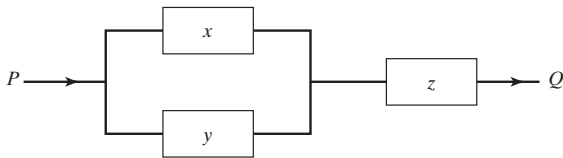


Figure 12. Diagram for Exercise 42.

the game will operate though component x is not operative; and let C be the event that the game will operate though component y is not operative. Using the notation in which $(0, 0, 1)$, for example, denotes that component z is operative but components x and y are not,

- (a) list the elements of the sample space S and also the elements of S corresponding to events A , B , and C ;
- (b) determine which pairs of events, A and B , A and C , or B and C , are mutually exclusive.

43. An experiment consists of rolling a die until a 3 appears. Describe the sample space and determine

- (a) how many elements of the sample space correspond to the event that the 3 appears on the k th roll of the die;
- (b) how many elements of the sample space correspond to the event that the 3 appears not later than the k th roll of the die.

44. Express symbolically the sample space S that consists of all the points (x, y) on or in the circle of radius 3 centered at the point $(2, -3)$.

45. If $S = \{x | 0 < x < 10\}$, $M = \{x | 3 < x \leq 8\}$, and $N = \{x | 5 < x < 10\}$, find

- (a) $M \cup N$; (b) $M \cap N$;
- (c) $M \cap N'$; (d) $M' \cup N$.

46. In Figure 13, L is the event that a driver has liability insurance and C is the event that she has collision insurance. Express in words what events are represented by regions 1, 2, 3, and 4.

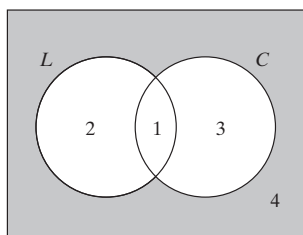


Figure 13. Venn diagram for Exercise 46.

47. With reference to Exercise 46 and Figure 13, what events are represented by

- (a) regions 1 and 2 together;
- (b) regions 2 and 4 together;

- (c) regions 1, 2, and 3 together;
- (d) regions 2, 3, and 4 together?

48. In Figure 14, E , T , and N are the events that a car brought to a garage needs an engine overhaul, transmission repairs, or new tires. Express in words the events represented by

- (a) region 1;
- (b) region 3;
- (c) region 7;
- (d) regions 1 and 4 together;
- (e) regions 2 and 5 together;
- (f) regions 3, 5, 6, and 8 together.

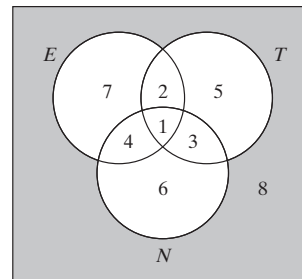


Figure 14. Venn diagram for Exercise 48.

49. With reference to Exercise 48 and Figure 14, list the region or combinations of regions representing the events that a car brought to the garage needs

- (a) transmission repairs, but neither an engine overhaul nor new tires;
- (b) an engine overhaul and transmission repairs;
- (c) transmission repairs or new tires, but not an engine overhaul;
- (d) new tires.

50. A market research organization claims that, among 500 shoppers interviewed, 308 regularly buy Product X , 266 regularly buy Product Y , 103 regularly buy both, and 59 buy neither on a regular basis. Using a Venn diagram and filling in the number of shoppers associated with the various regions, check whether the results of this study should be questioned.

51. In a group of 200 college students, 138 are enrolled in a course in psychology, 115 are enrolled in a course in sociology, and 91 are enrolled in both. How many of these students are not enrolled in either course? (*Hint:* Draw a suitable Venn diagram and fill in the numbers associated with the various regions.)

52. Among 120 visitors to Disneyland, 74 stayed for at least 3 hours, 86 spent at least \$20, 64 went on the Matterhorn ride, 60 stayed for at least 3 hours and spent at

least \$20, 52 stayed for at least 3 hours and went on the Matterhorn ride, 54 spent at least \$20 and went on the Matterhorn ride, and 48 stayed for at least 3 hours, spent at least \$20, and went on the Matterhorn ride. Drawing a Venn diagram with three circles (like that of Figure 4) and filling in the numbers associated with the various regions, find how many of the 120 visitors to Disneyland

- (a) stayed for at least 3 hours, spent at least \$20, but did not go on the Matterhorn ride;
 (b) went on the Matterhorn ride, but stayed less than 3 hours and spent less than \$20;
 (c) stayed less than 3 hours, spent at least \$20, but did not go on the Matterhorn ride.

SECS. 4–5

53. An experiment has five possible outcomes, A , B , C , D , and E , that are mutually exclusive. Check whether the following assignments of probabilities are permissible and explain your answers:

- (a) $P(A) = 0.20$, $P(B) = 0.20$, $P(C) = 0.20$, $P(D) = 0.20$, and $P(E) = 0.20$;
 (b) $P(A) = 0.21$, $P(B) = 0.26$, $P(C) = 0.58$, $P(D) = 0.01$, and $P(E) = 0.06$;
 (c) $P(A) = 0.18$, $P(B) = 0.19$, $P(C) = 0.20$, $P(D) = 0.21$, and $P(E) = 0.22$;
 (d) $P(A) = 0.10$, $P(B) = 0.30$, $P(C) = 0.10$, $P(D) = 0.60$, and $P(E) = -0.10$;
 (e) $P(A) = 0.23$, $P(B) = 0.12$, $P(C) = 0.05$, $P(D) = 0.50$, and $P(E) = 0.08$.

54. If A and B are mutually exclusive, $P(A) = 0.37$, and $P(B) = 0.44$, find

- (a) $P(A')$; (b) $P(B')$; (c) $P(A \cup B)$;
 (d) $P(A \cap B)$; (e) $P(A \cap B')$; (f) $P(A' \cap B')$.

55. Explain why there must be a mistake in each of the following statements:

- (a) The probability that Jean will pass the bar examination is 0.66 and the probability that she will not pass is -0.34 .
 (b) The probability that the home team will win an upcoming football game is 0.77, the probability that it will tie the game is 0.08, and the probability that it will win or tie the game is 0.95.
 (c) The probabilities that a secretary will make 0, 1, 2, 3, 4, or 5 or more mistakes in typing a report are, respectively, 0.12, 0.25, 0.36, 0.14, 0.09, and 0.07.
 (d) The probabilities that a bank will get 0, 1, 2, or 3 or more bad checks on any given day are, respectively, 0.08, 0.21, 0.29, and 0.40.

56. The probabilities that the serviceability of a new X-ray machine will be rated very difficult, difficult, average, easy, or very easy are, respectively, 0.12, 0.17, 0.34, 0.29, and 0.08. Find the probabilities that the serviceability of the machine will be rated

- (a) difficult or very difficult;
 (b) neither very difficult nor very easy;
 (c) average or worse;
 (d) average or better.

57. Suppose that each of the 30 points of the sample space of Exercise 40 is assigned the probability $\frac{1}{30}$. Find the probabilities that at a given moment

- (a) at least one of the station wagons is empty;
 (b) each of the two station wagons carries the same number of passengers;
 (c) the larger station wagon carries more passengers than the smaller station wagon;
 (d) together they carry at least six passengers.

58. A hat contains 20 white slips of paper numbered from 1 through 20, 10 red slips of paper numbered from 1 through 10, 40 yellow slips of paper numbered from 1 through 40, and 10 blue slips of paper numbered from 1 through 10. If these 80 slips of paper are thoroughly shuffled so that each slip has the same probability of being drawn, find the probabilities of drawing a slip of paper that is

- (a) blue or white;
 (b) numbered 1, 2, 3, 4, or 5;
 (c) red or yellow and also numbered 1, 2, 3, or 4;
 (d) numbered 5, 15, 25, or 35;
 (e) white and numbered higher than 12 or yellow and numbered higher than 26.

59. A police department needs new tires for its patrol cars and the probabilities are 0.15, 0.24, 0.03, 0.28, 0.22, and 0.08, respectively, that it will buy Uniroyal tires, Goodyear tires, Michelin tires, General tires, Goodrich tires, or Armstrong tires. Find the probabilities that it will buy

- (a) Goodyear or Goodrich tires;
 (b) Uniroyal, Michelin, or Goodrich tires;
 (c) Michelin or Armstrong tires;
 (d) Uniroyal, Michelin, General, or Goodrich tires.

60. Two cards are randomly drawn from a deck of 52 playing cards. Find the probability that both cards will be greater than 3 and less than 8.

61. Four candidates are seeking a vacancy on a school board. If A is twice as likely to be elected as B , and B and C are given about the same chance of being elected, while C is twice as likely to be elected as D , what are the probabilities that

- (a) C will win;
 (b) A will not win?

62. In a poker game, five cards are dealt at random from an ordinary deck of 52 playing cards. Find the probabilities of getting

Probability

- (a) two pairs (any two distinct face values occurring exactly twice);
 (b) four of a kind (four cards of equal face value).
- 63.** In a game of Yahtzee, five balanced dice are rolled simultaneously. Find the probabilities of getting
 (a) two pairs;
 (b) three of a kind;
 (c) a full house (three of a kind and a pair);
 (d) four of a kind.
- 64.** Explain on the basis of the various rules of Exercises 5 through 9 why there is a mistake in each of the following statements:
 (a) The probability that it will rain is 0.67, and the probability that it will rain or snow is 0.55.
 (b) The probability that a student will get a passing grade in English is 0.82, and the probability that she will get a passing grade in English and French is 0.86.
 (c) The probability that a person visiting the San Diego Zoo will see the giraffes is 0.72, the probability that he will see the bears is 0.84, and the probability that he will see both is 0.52.
- 65.** Among the 78 doctors on the staff of a hospital, 64 carry malpractice insurance, 36 are surgeons, and 34 of the surgeons carry malpractice insurance. If one of these doctors is chosen by lot to represent the hospital staff at an A.M.A. convention (that is, each doctor has a probability of $\frac{1}{78}$ of being selected), what is the probability that the one chosen is not a surgeon and does not carry malpractice insurance?
- 66.** A line segment of length l is divided by a point selected at random within the segment. What is the probability that it will divide the line segment in a ratio greater than 1:2? What is the probability that it will divide the segment exactly in half?
- 67.** A right triangle has the legs 3 and 4 units, respectively. Find the probability that a line segment, drawn at random parallel to the hypotenuse and contained entirely in the triangle, will divide the triangle so that the area between the line and the vertex opposite the hypotenuse will equal at least half the area of the triangle.
- 68.** For married couples living in a certain suburb, the probability that the husband will vote in a school board election is 0.21, the probability that the wife will vote in the election is 0.28, and the probability that they will both vote is 0.15. What is the probability that at least one of them will vote?
- 69.** Given $P(A) = 0.59$, $P(B) = 0.30$, and $P(A \cap B) = 0.21$, find
 (a) $P(A \cup B)$; (b) $P(A \cap B')$;
 (c) $P(A' \cup B')$; (d) $P(A' \cap B')$.
- 70.** At Roanoke College it is known that $\frac{1}{3}$ of the students live off campus. It is also known that $\frac{5}{9}$ of the students are from within the state of Virginia and that $\frac{3}{4}$ of the students are from out of state or live on campus. What is the probability that a student selected at random from Roanoke College is from out of state and lives on campus?
- 71.** A biology professor has two graduate assistants helping her with her research. The probability that the older of the two assistants will be absent on any given day is 0.08, the probability that the younger of the two will be absent on any given day is 0.05, and the probability that they will both be absent on any given day is 0.02. Find the probabilities that
 (a) either or both of the graduate assistants will be absent on any given day;
 (b) at least one of the two graduate assistants will not be absent on any given day;
 (c) only one of the two graduate assistants will be absent on any given day.
- 72.** Suppose that if a person visits Disneyland, the probability that he will go on the Jungle Cruise is 0.74, the probability that he will ride the Monorail is 0.70, the probability that he will go on the Matterhorn ride is 0.62, the probability that he will go on the Jungle Cruise and ride the Monorail is 0.52, the probability that he will go on the Jungle Cruise as well as the Matterhorn ride is 0.46, the probability that he will ride the Monorail and go on the Matterhorn ride is 0.44, and the probability that he will go on all three of these rides is 0.34. What is the probability that a person visiting Disneyland will go on at least one of these three rides?
- 73.** Suppose that if a person travels to Europe for the first time, the probability that he will see London is 0.70, the probability that he will see Paris is 0.64, the probability that he will see Rome is 0.58, the probability that he will see Amsterdam is 0.58, the probability that he will see London and Paris is 0.45, the probability that he will see London and Rome is 0.42, the probability that he will see London and Amsterdam is 0.41, the probability that he will see Paris and Rome is 0.35, the probability that he will see Paris and Amsterdam is 0.39, the probability that he will see Rome and Amsterdam is 0.32, the probability that he will see London, Paris, and Rome is 0.23, the probability that he will see London, Paris, and Amsterdam is 0.26, the probability that he will see London, Rome, and Amsterdam is 0.21, the probability that he will see Paris, Rome, and Amsterdam is 0.20, and the probability that he will see all four of these cities is 0.12. What is the probability that a person traveling to Europe for the first time will see at least one of these four cities? (*Hint:* Use the formula of Exercise 13.)

74. Use the formula of Exercise 15 to convert each of the following odds to probabilities:

(a) If three eggs are randomly chosen from a carton of 12 eggs of which 3 are cracked, the odds are 34 to 21 that at least one of them will be cracked.

(b) If a person has eight \$1 bills, five \$5 bills, and one \$20 bill, and randomly selects three of them, the odds are 11 to 2 that they will not all be \$1 bills.

(c) If we arbitrarily arrange the letters in the word “nest,” the odds are 5 to 1 that we will not get a meaningful word in the English language.

75. Use the definition of “odds” given in Exercise 15 to convert each of the following probabilities to odds:

(a) The probability that the last digit of a car’s license plate is a 2, 3, 4, 5, 6, or 7 is $\frac{6}{10}$.

(b) The probability of getting at least two heads in four flips of a balanced coin is $\frac{11}{16}$.

(c) The probability of rolling “7 or 11” with a pair of balanced dice is $\frac{2}{9}$.

SECS. 6–8

76. There are 90 applicants for a job with the news department of a television station. Some of them are college graduates and some are not; some of them have at least three years’ experience and some have not, with the exact breakdown being

	College graduates	Not college graduates
<i>At least three years’ experience</i>	18	9
<i>Less than three years’ experience</i>	36	27

If the order in which the applicants are interviewed by the station manager is random, G is the event that the first applicant interviewed is a college graduate, and T is the event that the first applicant interviewed has at least three years’ experience, determine each of the following probabilities directly from the entries and the row and column totals of the table:

(a) $P(G)$; **(b)** $P(T)$; **(c)** $P(G \cap T)$;

(d) $P(G' \cap T')$; **(e)** $P(T|G)$; **(f)** $P(G'|T')$.

77. Use the results of Exercise 76 to verify that

(a) $P(T|G) = \frac{P(G \cap T)}{P(G)}$;

(b) $P(G'|T') = \frac{P(G' \cap T')}{P(T')}$.

78. With reference to Exercise 65, what is the probability that the doctor chosen to represent the hospital staff at the convention carries malpractice insurance given that he or she is a surgeon?

79. With reference to Exercise 68, what is the probability that a husband will vote in the election given that his wife is going to vote?

80. With reference to Exercise 70, what is the probability that one of the students will be living on campus given that he or she is from out of state?

81. A bin contains 100 balls, of which 25 are red, 40 are white, and 35 are black. If two balls are selected from the bin without replacement, what is the probability that one will be red and one will be white?

82. If subjective probabilities are determined by the method suggested in Exercise 16, the third postulate of probability may not be satisfied. However, proponents of the subjective probability concept usually impose this postulate as a **consistency criterion**; in other words, they regard subjective probabilities that do not satisfy the postulate as inconsistent.

(a) A high school principal feels that the odds are 7 to 5 against her getting a \$1,000 raise and 11 to 1 against her getting a \$2,000 raise. Furthermore, she feels that it is an even-money bet that she will get one of these raises or the other. Discuss the consistency of the corresponding subjective probabilities.

(b) Asked about his political future, a party official replies that the odds are 2 to 1 that he will not run for the House of Representatives and 4 to 1 that he will not run for the Senate. Furthermore, he feels that the odds are 7 to 5 that he will run for one or the other. Are the corresponding probabilities consistent?

83. If we let $x =$ the number of spots facing up when a pair of dice is cast, then we can use the sample space S_2 of Example 2 to describe the outcomes of the experiment.

(a) Find the probability of each outcome in S_2 .

(b) Verify that the sum of these probabilities is 1.

84. There are two Porsches in a road race in Italy, and a reporter feels that the odds against their winning are 3 to 1 and 5 to 3. To be consistent (see Exercise 82), what odds should the reporter assign to the event that either car will win?

85. Using a computer program that can generate random integers on the interval $(0, 9)$ with equal probabilities, generate 1,000 such integers and use the frequency interpretation to estimate the probability that such a randomly chosen integer will have a value less than 1.

86. Using the method of Exercise 85, generate a second set of 1,000 random integers on $(0, 9)$. Estimate the probability that A : an integer selected at random from the first set will be less than 1 or B : an integer selected at random from the second set will be less than 1

(a) using the frequency interpretation of probabilities;

(b) using Theorem 7 and $P(A \cap B) = \frac{1}{81}$.

Probability

- 87.** It is felt that the probabilities are 0.20, 0.40, 0.30, and 0.10 that the basketball teams of four universities, T , U , V , and W , will win their conference championship. If university U is placed on probation and declared ineligible for the championship, what is the probability that university T will win the conference championship?
- 88.** With reference to Exercise 72, find the probabilities that a person who visits Disneyland will
- (a) ride the Monorail given that he will go on the Jungle Cruise;
 - (b) go on the Matterhorn ride given that he will go on the Jungle Cruise and ride the Monorail;
 - (c) not go on the Jungle Cruise given that he will ride the Monorail and/or go on the Matterhorn ride;
 - (d) go on the Matterhorn ride and the Jungle Cruise given that he will not ride the Monorail.
- (Hint: Draw a Venn diagram and fill in the probabilities associated with the various regions.)
- 89.** Crates of eggs are inspected for blood clots by randomly removing three eggs in succession and examining their contents. If all three eggs are good, the crate is shipped; otherwise it is rejected. What is the probability that a crate will be shipped if it contains 120 eggs, of which 10 have blood clots?
- 90.** The probability of surviving a certain transplant operation is 0.55. If a patient survives the operation, the probability that his or her body will reject the transplant within a month is 0.20. What is the probability of surviving both of these critical stages?
- 91.** Suppose that in Vancouver, B.C., the probability that a rainy fall day is followed by a rainy day is 0.80 and the probability that a sunny fall day is followed by a rainy day is 0.60. Find the probabilities that a rainy fall day is followed by
- (a) a rainy day, a sunny day, and another rainy day;
 - (b) two sunny days and then a rainy day;
 - (c) two rainy days and then two sunny days;
 - (d) rain two days later.
- [Hint: In part (c) use the formula of Exercise 19.]
- 92.** Use the formula of Exercise 19 to find the probability of randomly choosing (without replacement) four healthy guinea pigs from a cage containing 20 guinea pigs, of which 15 are healthy and 5 are diseased.
- 93.** A sharpshooter hits a target with probability 0.75. Assuming independence, find the probabilities of getting
- (a) a hit followed by two misses;
 - (b) two hits and a miss in any order.
- 94.** A balanced die is tossed twice. If A is the event that an even number comes up on the first toss, B is the event that an even number comes up on the second toss, and C is the event that both tosses result in the same number, are the events A , B , and C
- (a) pairwise independent;
 - (b) independent?
- 95.** A shipment of 1,000 parts contains 1 percent defective parts. Find the probability that
- (a) the first four parts chosen arbitrarily for inspection are nondefective;
 - (b) the first defective part found will be on the fourth inspection.
- 96.** A coin is loaded so that the probabilities of heads and tails are 0.52 and 0.48, respectively. If the coin is tossed three times, what are the probabilities of getting
- (a) all heads;
 - (b) two tails and a head in that order?
- 97.** If 5 of a company's 10 delivery trucks do not meet emission standards and 3 of them are chosen for inspection, what is the probability that none of the trucks chosen will meet emission standards?
- 98.** Medical records show that one out of 10 persons in a certain town has a thyroid deficiency. If 12 persons in this town are randomly chosen and tested, what is the probability that at least one of them will have a thyroid deficiency?
- 99.** If a person randomly picks 4 of the 15 gold coins a dealer has in stock, and 6 of the coins are counterfeits, what is the probability that the coins picked will all be counterfeits?
- 100.** A department store that bills its charge-account customers once a month has found that if a customer pays promptly one month, the probability is 0.90 that he or she will also pay promptly the next month; however, if a customer does not pay promptly one month, the probability that he or she will pay promptly the next month is only 0.40. (Assume that the probability of paying or not paying on any given month depends only on the outcome of the previous month.)
- (a) What is the probability that a customer who pays promptly one month will also pay promptly the next three months?
 - (b) What is the probability that a customer who does not pay promptly one month will also not pay promptly the next two months and then make a prompt payment the month after that?
- 101.** With reference to Figure 15, verify that events A , B , C , and D are independent. Note that the region representing A consists of two circles, and so do the regions representing B and C .

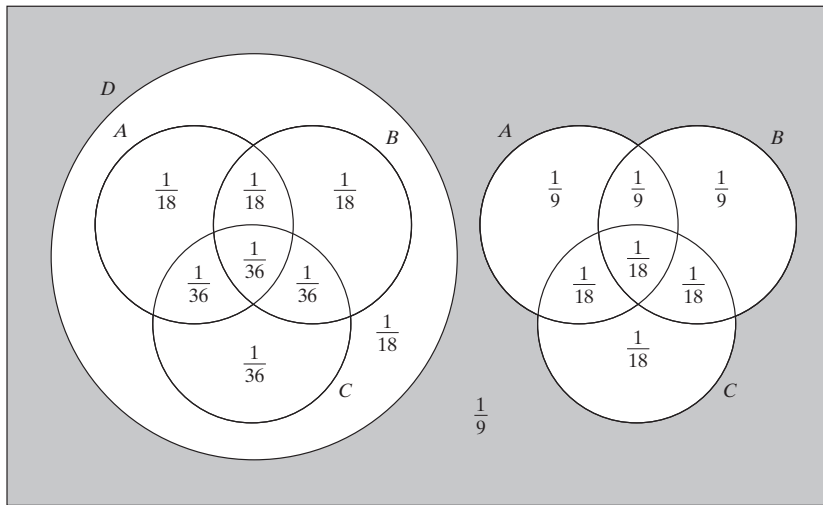


Figure 15. Diagram for Exercise 101.

102. At an electronics plant, it is known from past experience that the probability is 0.84 that a new worker who has attended the company's training program will meet the production quota, and that the corresponding probability is 0.49 for a new worker who has not attended the company's training program. If 70 percent of all new workers attend the training program, what is the probability that a new worker will meet the production quota?

103. It is known from experience that in a certain industry 60 percent of all labor-management disputes are over wages, 15 percent are over working conditions, and 25 percent are over fringe issues. Also, 45 percent of the disputes over wages are resolved without strikes, 70 percent of the disputes over working conditions are resolved without strikes, and 40 percent of the disputes over fringe issues are resolved without strikes. What is the probability that a labor-management dispute in this industry will be resolved without a strike?

104. In a T-maze, a rat is given food if it turns left and an electric shock if it turns right. On the first trial there is a 50–50 chance that a rat will turn either way; then, if it receives food on the first trial, the probability is 0.68 that it will turn left on the next trial, and if it receives a shock on the first trial, the probability is 0.84 that it will turn left on the next trial. What is the probability that a rat will turn left on the second trial?

105. With reference to Exercise 103, what is the probability that if a labor-management dispute in this industry is resolved without a strike, it was over wages?

106. The probability that a one-car accident is due to faulty brakes is 0.04, the probability that a one-car accident is correctly attributed to faulty brakes is 0.82, and

the probability that a one-car accident is incorrectly attributed to faulty brakes is 0.03. What is the probability that

- (a) a one-car accident will be attributed to faulty brakes;
- (b) a one-car accident attributed to faulty brakes was actually due to faulty brakes?

107. With reference to Example 25, suppose that we discover later that the job was completed on time. What is the probability that there had been a strike?

108. In a certain community, 8 percent of all adults over 50 have diabetes. If a health service in this community correctly diagnoses 95 percent of all persons with diabetes as having the disease and incorrectly diagnoses 2 percent of all persons without diabetes as having the disease, find the probabilities that

- (a) the community health service will diagnose an adult over 50 as having diabetes;
- (b) a person over 50 diagnosed by the health service as having diabetes actually has the disease.

109. An explosion at a construction site could have occurred as the result of static electricity, malfunctioning of equipment, carelessness, or sabotage. Interviews with construction engineers analyzing the risks involved led to the estimates that such an explosion would occur with probability 0.25 as a result of static electricity, 0.20 as a result of malfunctioning of equipment, 0.40 as a result of carelessness, and 0.75 as a result of sabotage. It is also felt that the prior probabilities of the four causes of the explosion are 0.20, 0.40, 0.25, and 0.15. Based on all this information, what is

- (a) the most likely cause of the explosion;
- (b) the least likely cause of the explosion?

110. A mail-order house employs three stock clerks, U , V , and W , who pull items from shelves and assemble them for subsequent verification and packaging. U makes a mistake in an order (gets a wrong item or the wrong quantity) one time in a hundred, V makes a mistake in an order five times in a hundred, and W makes a mistake in an order three times in a hundred. If U , V , and W fill, respectively, 30, 40, and 30 percent of all orders, what are the probabilities that

(a) a mistake will be made in an order;

(b) if a mistake is made in an order, the order was filled by U ;

(c) if a mistake is made in an order, the order was filled by V ?

111. An art dealer receives a shipment of five old paintings from abroad, and, on the basis of past experience, she feels that the probabilities are, respectively, 0.76, 0.09, 0.02, 0.01, 0.02, and 0.10 that 0, 1, 2, 3, 4, or all 5 of them are forgeries. Since the cost of authentication is fairly high, she decides to select one of the five paintings at random and send it away for authentication. If it turns out that this painting is a forgery, what probability should she now assign to the possibility that all the other paintings are also forgeries?

112. To get answers to sensitive questions, we sometimes use a method called the **randomized response technique**. Suppose, for instance, that we want to determine what percentage of the students at a large university smoke marijuana. We construct 20 flash cards, write “I smoke marijuana at least once a week” on 12 of the cards, where 12 is an arbitrary choice, and “I do not smoke marijuana at least once a week” on the others. Then, we let each student (in the sample interviewed) select one of the 20 cards at random and respond “yes” or “no” without divulging the question.

(a) Establish a relationship between $P(Y)$, the probability that a student will give a “yes” response, and $P(M)$, the probability that a student randomly selected at that university smokes marijuana at least once a week.

(b) If 106 of 250 students answered “yes” under these conditions, use the result of part (a) and $\frac{106}{250}$ as an estimate of $P(Y)$ to estimate $P(M)$.

SEC. 9

113. A series system consists of three components, each having the reliability 0.95, and three components, each having the reliability 0.99. Find the reliability of the system.

114. Find the reliability of a series systems having five components with reliabilities 0.995, 0.990, 0.992, 0.995, and 0.998, respectively.

115. What must be the reliability of each component in a series system consisting of six components that must have a system reliability of 0.95?

116. Referring to Exercise 115, suppose now that there are 10 components, and the system reliability must be 0.90.

117. Suppose a system consists of four components, connected in parallel, having the reliabilities 0.8, 0.7, 0.7, and 0.65, respectively. Find the system reliability.

118. Referring to Exercise 117, suppose now that the system has five components with reliabilities 0.85, 0.80, 0.65, 0.60, and 0.70, respectively. Find the system reliability.

119. A system consists of two components having the reliabilities 0.95 and 0.90, connected in series to two parallel subsystems, the first containing four components, each having the reliability 0.60, and the second containing two components, each having the reliability 0.75. Find the system reliability.

120. A series system consists of two components having the reliabilities 0.98 and 0.99, respectively, connected to a parallel subsystem containing five components having the reliabilities 0.75, 0.60, 0.65, 0.70, and 0.60, respectively. Find the system reliability.

References

Among the numerous textbooks on probability theory published in recent years, one of the most popular is

FELLER, W., *An Introduction to Probability Theory and Its Applications*, Vol. I, 3rd ed. New York: John Wiley & Sons, Inc., 1968.

More elementary treatments may be found in

BARR, D. R., and ZEHNA, P. W., *Probability: Modeling Uncertainty*. Reading, Mass.: Addison-Wesley Publishing Company, Inc., 1983,

DRAPER, N. R., and LAWRENCE, W. E., *Probability: An Introductory Course*. Chicago: Markam Publishing Company, 1970,

FREUND, J. E., *Introduction to Probability*. New York: Dover Publications, Inc., 1993 Reprint,

GOLDBERG, S., *Probability—An Introduction*. Mineola, N.Y.: Dover Publications, Inc. (republishing of 1960 edition),

HODGES, J. L., and LEHMANN, E. L., *Elements of Finite Probability*. San Francisco: Holden Day, Inc., 1965,

NOSAL, M., *Basic Probability and Applications*. Philadelphia: W. B. Saunders Company, 1977.

More advanced treatments are given in many texts, for instance,

Probability

HOEL, P., PORT, S. C., and STONE, C. J., *Introduction to Probability Theory*. Boston: Houghton Mifflin Company, 1971,
 KHAZANIE, R., *Basic Probability Theory and Applications*. Pacific Palisades, Calif.: Goodyear Publishing Company, Inc., 1976,
 PARZEN, E., *Modern Probability Theory and Its Applications*. New York: John Wiley & Sons, Inc., 1960,
 ROSS, S., *A First Course in Probability*, 3rd ed. New York: Macmillan Publishing Company, 1988.

SOLOMON, F., *Probability and Stochastic Processes*. Upper Saddle River, N.J.: Prentice Hall, 1987.

For more on reliability and related topics, see
 JOHNSON, R. A., *Miller and Freund's Probability and Statistics for Engineers*. Upper Saddle River, N.J.: Prentice Hall, 2000,
 MILLER, I. and MILLER, M., *Statistical Methods for Quality with Applications to Engineering and Management*. Upper Saddle River, N.J.: Prentice Hall, 1995.

Answers to Odd-Numbered Exercises

- 35 (a) {6, 8, 9}; (b) {8}; (c) {1, 2, 3, 4, 5, 8}; (d) {1, 5}; (e) {2, 4, 8}; (f) \emptyset .
 37 (a) {Car 5, Car 6, Car 7, Car 8}; (b) {Car 2, Car 4, Car 5, Car 7}; (c) {Car 1, Car 8}; (d) {Car 3, Car 4, Car 7, Car 8}.
 39 (a) The house has fewer than three baths. (b) The house does not have a fireplace. (c) The house does not cost more than \$200,000. (d) The house is not new. (e) The house has three or more baths and a fireplace. (f) The house has three or more baths and costs more than \$200,000. (g) The house costs more than \$200,000 but has no fireplace. (h) The house is new or costs more than \$200,000. (i) The house is new or costs at most \$200,000. (j) The house has three or more baths and/or a fireplace. (k) The house has three or more baths and/or costs more than \$200,000. (l) The house is new and costs more than \$200,000.
 41 (a) (H,1), (H,2), (H,3), (H,4), (H,5), (H,6), (T,H,H), (T,H,T), (T,T,H), and (T,T,T); (b) (H,1), (H,2), (H,3), (H,4), (H,5), (H,6), (T,H,T), and (T,T,H); (c) (H,5), (H,6), (T,H,T), (T,T,H), and (T,T,T).
 43 (a) 5^{k-1} ; (b) $\frac{5^k - 1}{4}$.
 45 (a) $(x|3 < x < 10)$; (b) $(x|15 < x \leq 8)$; (c) $(x|3 < x \leq 5)$; (d) $(x|0 < x \leq 3)$ or $(5 < x < 10)$.
 47 (a) The event that a driver has liability insurance. (b) The event that a driver does not have collision insurance. (c) The event that a driver has liability insurance or collision insurance, but not both. (d) The event that a driver does not have both kinds of insurance.
 49 (a) Region 5; (b) regions 1 and 2 together; (c) regions 3, 5, and 6 together; (d) regions 1, 3, 4, and 6 together.
 51 38.
 53 (a) Permissible; (b) not permissible because the sum of the probabilities exceeds 1; (c) permissible; (d) not permissible because $P(E)$ is negative; (e) not permissible because the sum of the probabilities is less than 1.
 55 (a) The probability that she cannot pass cannot be negative. (b) $0.77 + 0.08 = 0.85 \neq 0.95$; (c) $0.12 + 0.25 + 0.36 + 0.14 + 0.09 + 0.07 = 1.03 > 1$; (d) $0.08 + 0.21 + 0.29 + 0.40 = 0.98 < 1$.

- 57 (a) $\frac{1}{3}$; (b) $\frac{1}{6}$; (c) $\frac{1}{2}$; (d) $\frac{1}{3}$.
 59 (a) 0.46; (b) 0.40; (c) 0.11 (d) 0.68.
 61 (a) $\frac{2}{9}$; (b) $\frac{5}{9}$.
 63 (a) $\frac{25}{108}$; (b) $\frac{25}{162}$; (c) $\frac{25}{648}$; (d) $\frac{25}{1296}$.
 65 $\frac{2}{13}$.
 67 $1 - \frac{\sqrt{2}}{2}$.
 69 (a) 0.68; (b) 0.38; (c) 0.79; (d) 0.32.
 71 (a) 0.11; (b) 0.98; (c) 0.09.
 73 0.94.
 75 (a) 3 to 2; (b) 11 to 5; (c) 7 to 2 against it.
 77 (a) $\frac{1}{3}$; (b) $\frac{3}{7}$.
 79 $\frac{15}{28}$.
 81 (a) 0.2; (b) $\frac{20}{99}$.
 83

Outcome	2	3	4	5	6	7	8	9	10	11	12
Probability	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{12}$	$\frac{1}{9}$	$\frac{5}{36}$	$\frac{1}{6}$	$\frac{5}{36}$	$\frac{1}{9}$	$\frac{1}{12}$	$\frac{1}{18}$	$\frac{1}{36}$

 87 $\frac{1}{3}$.
 89 0.7685.
 91 (a) 0.096; (b) 0.048; (c) 0.0512; (d) 0.76.
 93 (a) $\frac{3}{64}$; (b) $\frac{27}{64}$.
 95 (a) Required probability = 0.9606; exact probability = 0.9605; (b) required probability = 0.0097 (assuming independence); exact probability = 0.0097.
 97 $\frac{1}{12}$.
 99 $\frac{1}{91}$.
 103 0.475.
 105 0.5684.
 107 0.3818.
 109 (a) Most likely cause is sabotage ($P = 0.3285$); (b) least likely cause is static electricity ($P = 0.1460$).
 111 0.6757.
 113 0.832.
 115 0.991.
 117 0.9937.
 119 0.781.

PROBABILITY DISTRIBUTIONS AND PROBABILITY DENSITIES

- | | |
|---------------------------------|------------------------------|
| 1 Random Variables | 5 Multivariate Distributions |
| 2 Probability Distributions | 6 Marginal Distributions |
| 3 Continuous Random Variables | 7 Conditional Distributions |
| 4 Probability Density Functions | 8 The Theory in Practice |

I Random Variables

In most applied problems involving probabilities we are interested only in a particular aspect (or in two or a few particular aspects) of the outcomes of experiments. For instance, when we roll a pair of dice we are usually interested only in the total, and not in the outcome for each die; when we interview a randomly chosen married couple we may be interested in the size of their family and in their joint income, but not in the number of years they have been married or their total assets; and when we sample mass-produced light bulbs we may be interested in their durability or their brightness, but not in their price.

In each of these examples we are interested in numbers that are associated with the outcomes of chance experiments, that is, in the values taken on by **random variables**. In the language of probability and statistics, the total we roll with a pair of dice is a random variable, the size of the family of a randomly chosen married couple and their joint income are random variables, and so are the durability and the brightness of a light bulb randomly picked for inspection.

To be more explicit, consider Figure 1, which pictures the sample space for an experiment in which we roll a pair of dice, and let us assume that each of the 36 possible outcomes has the probability $\frac{1}{36}$. Note, however, that in Figure 1 we have attached a number to each point: For instance, we attached the number 2 to the point (1, 1), the number 6 to the point (1, 5), the number 8 to the point (6, 2), the number 11 to the point (5, 6), and so forth. Evidently, we associated with each point the value of a random variable, that is, the corresponding total rolled with the pair of dice.

Since “associating a number with each point (element) of a sample space” is merely another way of saying that we are “defining a function over the points of a sample space,” let us now make the following definition.

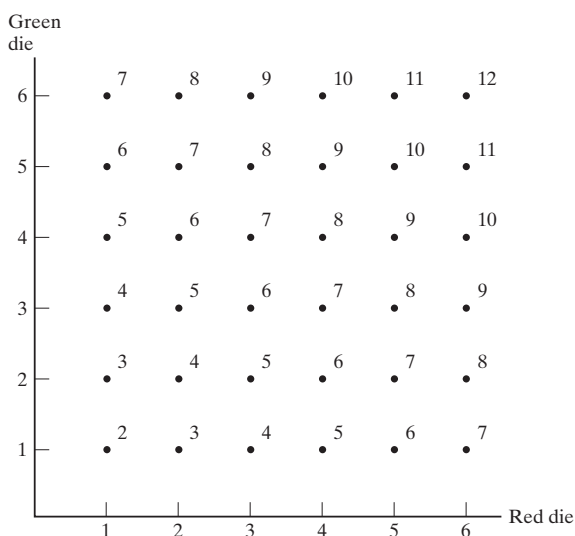


Figure 1. The total number of points rolled with a pair of dice.

DEFINITION 1. RANDOM VARIABLE. *If S is a sample space with a probability measure and X is a real-valued function defined over the elements of S , then X is called a random variable.*[†]

In this chapter we shall denote random variables by capital letters and their values by the corresponding lowercase letters; for instance, we shall write x to denote a value of the random variable X .

With reference to the preceding example and Figure 1, observe that the random variable X takes on the value 9, and we write $X = 9$ for the subset

$$\{(6, 3), (5, 4), (4, 5), (3, 6)\}$$

of the sample space S . Thus, $X = 9$ is to be interpreted as the set of elements of S for which the total is 9 and, more generally, $X = x$ is to be interpreted as the set of elements of the sample space for which the random variable X takes on the value x . This may seem confusing, but it reminds one of mathematicians who say “ $f(x)$ is a function of x ” instead of “ $f(x)$ is the value of a function at x .”

EXAMPLE 1

Two socks are selected at random and removed in succession from a drawer containing five brown socks and three green socks. List the elements of the sample space, the corresponding probabilities, and the corresponding values w of the random variable W , where W is the number of brown socks selected.

[†]Instead of “random variable,” the terms “chance variable,” “stochastic variable,” and “variate” are also used in some books.

Solution

If B and G stand for brown and green, the probabilities for BB , BG , GB , and GG are, respectively, $\frac{5}{8} \cdot \frac{4}{7} = \frac{5}{14}$, $\frac{5}{8} \cdot \frac{3}{7} = \frac{15}{56}$, $\frac{3}{8} \cdot \frac{5}{7} = \frac{15}{56}$, and $\frac{3}{8} \cdot \frac{2}{7} = \frac{3}{28}$, and the results are shown in the following table:

<i>Element of sample space</i>	<i>Probability</i>	<i>w</i>
BB	$\frac{5}{14}$	2
BG	$\frac{15}{56}$	1
GB	$\frac{15}{56}$	1
GG	$\frac{3}{28}$	0

Also, we can write $P(W = 2) = \frac{5}{14}$, for example, for the probability of the event that the random variable W will take on the value 2.

EXAMPLE 2

A balanced coin is tossed four times. List the elements of the sample space that are presumed to be equally likely, as this is what we mean by a coin being balanced, and the corresponding values x of the random variable X , the total number of heads.

Solution

If H and T stand for heads and tails, the results are as shown in the following table:

<i>Element of sample space</i>	<i>Probability</i>	<i>x</i>
HHHH	$\frac{1}{16}$	4
HHHT	$\frac{1}{16}$	3
HHTH	$\frac{1}{16}$	3
HTHH	$\frac{1}{16}$	3
THHH	$\frac{1}{16}$	3
HHTT	$\frac{1}{16}$	2
HTHT	$\frac{1}{16}$	2

<i>Element of sample space</i>	<i>Probability</i>	<i>x</i>
HTTH	$\frac{1}{16}$	2
THHT	$\frac{1}{16}$	2
THTH	$\frac{1}{16}$	2
TTHH	$\frac{1}{16}$	2
HTTT	$\frac{1}{16}$	1
THTT	$\frac{1}{16}$	1
TTHT	$\frac{1}{16}$	1
TTTH	$\frac{1}{16}$	1
TTTT	$\frac{1}{16}$	0

Thus, we can write $P(X = 3) = \frac{4}{16}$, for example, for the probability of the event that the random variable X will take on the value 3.

The fact that Definition 1 is limited to real-valued functions does not impose any restrictions. If the numbers we want to assign to the outcomes of an experiment are complex numbers, we can always look upon the real and the imaginary parts separately as values taken on by two random variables. Also, if we want to describe the outcomes of an experiment quantitatively, say, by giving the color of a person's hair, we can arbitrarily make the descriptions real-valued by coding the various colors, perhaps by representing them with the numbers 1, 2, 3, and so on.

In all of the examples of this section we have limited our discussion to discrete sample spaces, and hence to **discrete random variables**, namely, random variables whose range is finite or countably infinite. Continuous random variables defined over continuous sample spaces will be taken up in Section 3.

2 Probability Distributions

As we already saw in Examples 1 and 2, the probability measure defined over a discrete sample space automatically provides the probabilities that a random variable will take on any given value within its range.

For instance, having assigned the probability $\frac{1}{36}$ to each element of the sample space of Figure 1, we immediately find that the random variable X , the total rolled with the pair of dice, takes on the value 9 with probability $\frac{4}{36}$; as described in Section 1, $X = 9$ contains four of the equally likely elements of the sample space. The probabilities associated with all possible values of X are shown in the following table:

Probability Distributions and Probability Densities

x	$P(X = x)$
2	$\frac{1}{36}$
3	$\frac{2}{36}$
4	$\frac{3}{36}$
5	$\frac{4}{36}$
6	$\frac{5}{36}$
7	$\frac{6}{36}$
8	$\frac{5}{36}$
9	$\frac{4}{36}$
10	$\frac{3}{36}$
11	$\frac{2}{36}$
12	$\frac{1}{36}$

Instead of displaying the probabilities associated with the values of a random variable in a table, as we did in the preceding illustration, it is usually preferable to give a formula, that is, to express the probabilities by means of a function such that its values, $f(x)$, equal $P(X = x)$ for each x within the range of the random variable X . For instance, for the total rolled with a pair of dice we could write

$$f(x) = \frac{6 - |x - 7|}{36} \quad \text{for } x = 2, 3, \dots, 12$$

as can easily be verified by substitution. Clearly,

$$\begin{aligned}
 f(2) &= \frac{6 - |2 - 7|}{36} = \frac{6 - 5}{36} = \frac{1}{36} \\
 f(3) &= \frac{6 - |3 - 7|}{36} = \frac{6 - 4}{36} = \frac{2}{36} \\
 &\dots\dots\dots \\
 f(12) &= \frac{6 - |12 - 7|}{36} = \frac{6 - 5}{36} = \frac{1}{36}
 \end{aligned}$$

and all these values agree with the ones shown in the preceding table.

DEFINITION 2. PROBABILITY DISTRIBUTION. *If X is a discrete random variable, the function given by $f(x) = P(X = x)$ for each x within the range of X is called the **probability distribution** of X .*

Based on the postulates of probability, we obtain the following theorem.

THEOREM 1. A function can serve as the probability distribution of a discrete random variable X if and only if its values, $f(x)$, satisfy the conditions

1. $f(x) \geq 0$ for each value within its domain;
2. $\sum_x f(x) = 1$, where the summation extends over all the values within its domain.

EXAMPLE 3

Find a formula for the probability distribution of the total number of heads obtained in four tosses of a balanced coin.

Solution

Based on the probabilities in the table, we find that $P(X = 0) = \frac{1}{16}$, $P(X = 1) = \frac{4}{16}$, $P(X = 2) = \frac{6}{16}$, $P(X = 3) = \frac{4}{16}$, and $P(X = 4) = \frac{1}{16}$. Observing that the numerators of these five fractions, 1, 4, 6, 4, and 1, are the binomial coefficients $\binom{4}{0}$, $\binom{4}{1}$, $\binom{4}{2}$, $\binom{4}{3}$, and $\binom{4}{4}$, we find that the formula for the probability distribution can be written as

$$f(x) = \frac{\binom{4}{x}}{16} \quad \text{for } x = 0, 1, 2, 3, 4$$

EXAMPLE 4

Check whether the function given by

$$f(x) = \frac{x+2}{25} \quad \text{for } x = 1, 2, 3, 4, 5$$

can serve as the probability distribution of a discrete random variable.

Solution

Substituting the different values of x , we get $f(1) = \frac{3}{25}$, $f(2) = \frac{4}{25}$, $f(3) = \frac{5}{25}$, $f(4) = \frac{6}{25}$, and $f(5) = \frac{7}{25}$. Since these values are all nonnegative, the first condition of Theorem 1 is satisfied, and since

$$\begin{aligned} f(1) + f(2) + f(3) + f(4) + f(5) &= \frac{3}{25} + \frac{4}{25} + \frac{5}{25} + \frac{6}{25} + \frac{7}{25} \\ &= 1 \end{aligned}$$

the second condition of Theorem 1 is satisfied. Thus, the given function can serve as the probability distribution of a random variable having the range $\{1, 2, 3, 4, 5\}$. Of course, whether any given random variable actually has this probability distribution is an entirely different matter.

In some problems it is desirable to present probability distributions graphically, and two kinds of graphical presentations used for this purpose are shown in Figures 2 and 3. The one shown in Figure 2, called a **probability histogram**, represents the probability distribution of Example 3. The height of each rectangle equals

Probability Distributions and Probability Densities

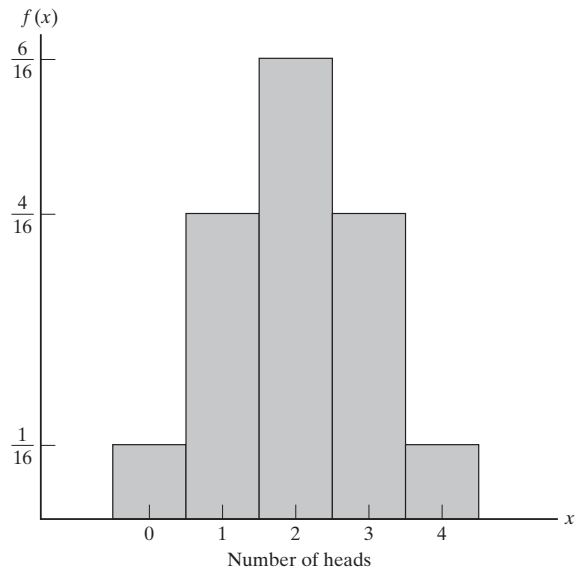


Figure 2. Probability histogram.

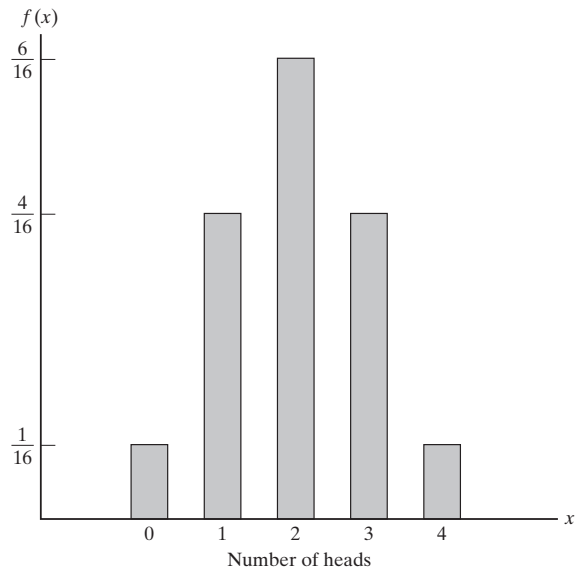


Figure 3. Bar chart.

the probability that X takes on the value that corresponds to the midpoint of its base. By representing 0 with the interval from -0.5 to 0.5 , 1 with the interval from 0.5 to 1.5 , \dots , and 4 with the interval from 3.5 to 4.5 , we are, so to speak, “spreading” the values of the given discrete random variable over a continuous scale.

Since each rectangle of the probability histogram of Figure 2 has unit width, we could have said that the *areas* of the rectangles, rather than their heights, equal the

corresponding probabilities. There are certain advantages to identifying the areas of the rectangles with the probabilities, for instance, when we wish to approximate the graph of a discrete probability distribution with a continuous curve. This can be done even when the rectangles of a probability histogram do not all have unit width by adjusting the heights of the rectangles or by modifying the vertical scale.

The graph of Figure 3 is called a **bar chart**, but it is also referred to as a histogram. As in Figure 2, the height of each rectangle, or bar, equals the probability of the corresponding value of the random variable, but there is no pretense of having a continuous horizontal scale. Sometimes, as shown in Figure 4, we use lines (rectangles with no width) instead of the rectangles, but we still refer to the graphs as probability histograms.

In this chapter, histograms and bar charts are used mainly in descriptive statistics to convey visually the information provided by a probability distribution or a distribution of actual data (see Section 8).

There are many problems in which it is of interest to know the probability that the value of a random variable is less than or equal to some real number x . Thus, let us write the probability that X takes on a value less than or equal to x as $F(x) = P(X \leq x)$ and refer to this function defined for all real numbers x as the **distribution function**, or the **cumulative distribution**, of X .

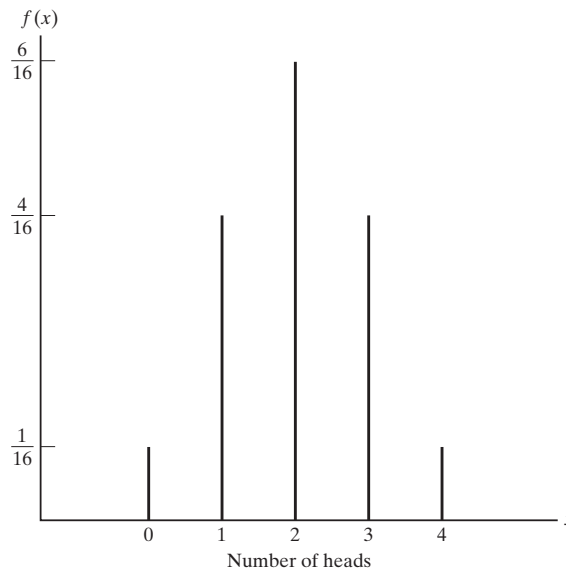


Figure 4. Probability histogram.

DEFINITION 3. DISTRIBUTION FUNCTION. If X is a discrete random variable, the function given by

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t) \quad \text{for } -\infty < x < \infty$$

where $f(t)$ is the value of the probability distribution of X at t , is called the **distribution function**, or the **cumulative distribution** of X .

Based on the postulates of probability and some of their immediate consequences, we obtain the following theorem.

THEOREM 2. The values $F(x)$ of the distribution function of a discrete random variable X satisfy the conditions

1. $F(-\infty) = 0$ and $F(\infty) = 1$;
2. if $a < b$, then $F(a) \leq F(b)$ for any real numbers a and b .

If we are given the probability distribution of a discrete random variable, the corresponding distribution function is generally easy to find.

EXAMPLE 5

Find the distribution function of the total number of heads obtained in four tosses of a balanced coin.

Solution

Given $f(0) = \frac{1}{16}, f(1) = \frac{4}{16}, f(2) = \frac{6}{16}, f(3) = \frac{4}{16}$, and $f(4) = \frac{1}{16}$ from Example 3, it follows that

$$\begin{aligned}
 F(0) &= f(0) = \frac{1}{16} \\
 F(1) &= f(0) + f(1) = \frac{5}{16} \\
 F(2) &= f(0) + f(1) + f(2) = \frac{11}{16} \\
 F(3) &= f(0) + f(1) + f(2) + f(3) = \frac{15}{16} \\
 F(4) &= f(0) + f(1) + f(2) + f(3) + f(4) = 1
 \end{aligned}$$

Hence, the distribution function is given by

$$F(x) = \begin{cases} 0 & \text{for } x < 0 \\ \frac{1}{16} & \text{for } 0 \leq x < 1 \\ \frac{5}{16} & \text{for } 1 \leq x < 2 \\ \frac{11}{16} & \text{for } 2 \leq x < 3 \\ \frac{15}{16} & \text{for } 3 \leq x < 4 \\ 1 & \text{for } x \geq 4 \end{cases}$$

Observe that this distribution function is defined not only for the values taken on by the given random variable, but for all real numbers. For instance, we can write $F(1.7) = \frac{5}{16}$ and $F(100) = 1$, although the probabilities of getting “at most 1.7 heads” or “at most 100 heads” in four tosses of a balanced coin may not be of any real significance.

EXAMPLE 6

Find the distribution function of the random variable W of Example 1 and plot its graph.

Solution

Based on the probabilities given in the table in Section 1, we can write $f(0) = \frac{3}{28}$, $f(1) = \frac{15}{56} + \frac{15}{56} = \frac{15}{28}$, and $f(2) = \frac{5}{14}$, so that

$$\begin{aligned}
 F(0) &= f(0) = \frac{3}{28} \\
 F(1) &= f(0) + f(1) = \frac{9}{14} \\
 F(2) &= f(0) + f(1) + f(2) = 1
 \end{aligned}$$

Hence, the distribution function of W is given by

$$F(w) = \begin{cases} 0 & \text{for } w < 0 \\ \frac{3}{28} & \text{for } 0 \leq w < 1 \\ \frac{9}{14} & \text{for } 1 \leq w < 2 \\ 1 & \text{for } w \geq 2 \end{cases}$$

The graph of this distribution function, shown in Figure 5, was obtained by first plotting the points $(w, F(w))$ for $w = 0, 1,$ and 2 and then completing the step function as indicated. Note that at all points of discontinuity the distribution function takes on the greater of the two values, as indicated by the heavy dots in Figure 5.

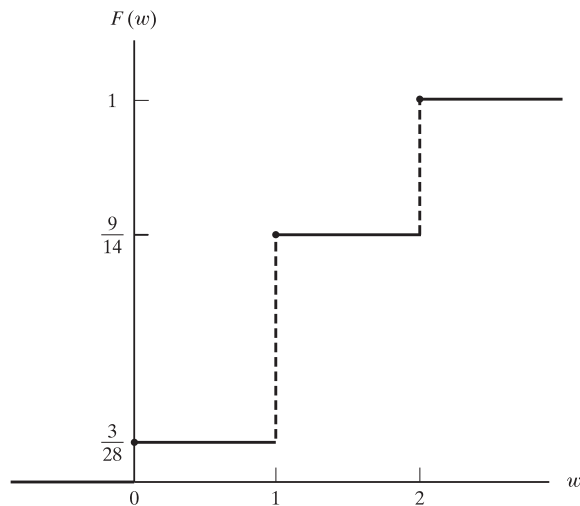


Figure 5. Graph of the distribution function of Example 6.

We can also reverse the process illustrated in the two preceding examples, that is, obtain values of the probability distribution of a random variable from its distribution function. To this end, we use the following result.

THEOREM 3. If the range of a random variable X consists of the values $x_1 < x_2 < x_3 < \dots < x_n$, then $f(x_1) = F(x_1)$ and

$$f(x_i) = F(x_i) - F(x_{i-1}) \quad \text{for } i = 2, 3, \dots, n$$

EXAMPLE 7

If the distribution function of X is given by

$$F(x) = \begin{cases} 0 & \text{for } x < 2 \\ \frac{1}{36} & \text{for } 2 \leq x < 3 \\ \frac{3}{36} & \text{for } 3 \leq x < 4 \\ \frac{6}{36} & \text{for } 4 \leq x < 5 \\ \frac{10}{36} & \text{for } 5 \leq x < 6 \\ \frac{15}{36} & \text{for } 6 \leq x < 7 \\ \frac{21}{36} & \text{for } 7 \leq x < 8 \\ \frac{26}{36} & \text{for } 8 \leq x < 9 \\ \frac{30}{36} & \text{for } 9 \leq x < 10 \\ \frac{33}{36} & \text{for } 10 \leq x < 11 \\ \frac{35}{36} & \text{for } 11 \leq x < 12 \\ 1 & \text{for } x \geq 12 \end{cases}$$

find the probability distribution of this random variable.

Solution

Making use of Theorem 3, we get $f(2) = \frac{1}{36}$, $f(3) = \frac{3}{36} - \frac{1}{36} = \frac{2}{36}$, $f(4) = \frac{6}{36} - \frac{3}{36} = \frac{3}{36}$, $f(5) = \frac{10}{36} - \frac{6}{36} = \frac{4}{36}$, \dots , $f(12) = 1 - \frac{35}{36} = \frac{1}{36}$, and comparison with the probabilities in the table in Section 2 reveals that the random variable with which we are concerned here is the total number of points rolled with a pair of dice.

In the remainder of this chapter we will be concerned with continuous random variables and their distributions and with problems relating to the simultaneous occurrence of the values of two or more random variables.

Exercises

1. For each of the following, determine whether the given values can serve as the values of a probability distribution of a random variable with the range $x = 1, 2, 3,$ and 4 :

(a) $f(1) = 0.25, f(2) = 0.75, f(3) = 0.25,$ and $f(4) = -0.25$;

(b) $f(1) = 0.15, f(2) = 0.27, f(3) = 0.29,$ and $f(4) = 0.29$;

(c) $f(1) = \frac{1}{19}, f(2) = \frac{10}{19}, f(3) = \frac{2}{19},$ and $f(4) = \frac{5}{19}$.

2. For each of the following, determine whether the given function can serve as the probability distribution of a random variable with the given range:

(a) $f(x) = \frac{x-2}{5}$ for $x = 1, 2, 3, 4, 5$;

(b) $f(x) = \frac{x^2}{30}$ for $x = 0, 1, 2, 3, 4$;

(c) $f(x) = \frac{1}{5}$ for $x = 0, 1, 2, 3, 4, 5$.

3. Verify that $f(x) = \frac{2x}{k(k+1)}$ for $x = 1, 2, 3, \dots, k$ can serve as the probability distribution of a random variable with the given range.

4. For each of the following, determine c so that the function can serve as the probability distribution of a random variable with the given range:

(a) $f(x) = cx$ for $x = 1, 2, 3, 4, 5$;

(b) $f(x) = c \binom{5}{x}$ for $x = 0, 1, 2, 3, 4, 5$;

(c) $f(x) = cx^2$ for $x = 1, 2, 3, \dots, k$;

(d) $f(x) = c \left(\frac{1}{4}\right)^x$ for $x = 1, 2, 3, \dots$

5. For what values of k can

$$f(x) = (1 - k)k^x$$

serve as the values of the probability distribution of a random variable with the countably infinite range $x = 0, 1, 2, \dots$?

6. Show that there are no values of c such that

$$f(x) = \frac{c}{x}$$

can serve as the values of the probability distribution of a random variable with the countably infinite range $x = 1, 2, 3, \dots$

7. Construct a probability histogram for each of the following probability distributions:

(a) $f(x) = \frac{\binom{2}{x} \binom{4}{3-x}}{\binom{6}{3}}$ for $x = 0, 1, 2$;

(b) $f(x) = \binom{5}{x} \left(\frac{1}{5}\right)^x \left(\frac{4}{5}\right)^{5-x}$ for $x = 0, 1, 2, 3, 4, 5$.

8. Prove Theorem 2.

9. For each of the following, determine whether the given values can serve as the values of a distribution function of a random variable with the range $x = 1, 2, 3,$ and 4 :

(a) $F(1) = 0.3, F(2) = 0.5, F(3) = 0.8,$ and $F(4) = 1.2$;

(b) $F(1) = 0.5, F(2) = 0.4, F(3) = 0.7,$ and $F(4) = 1.0$;

(c) $F(1) = 0.25, F(2) = 0.61, F(3) = 0.83,$ and $F(4) = 1.0$.

10. Find the distribution function of the random variable of part (a) of Exercise 7 and plot its graph.

11. If X has the distribution function

$$F(x) = \begin{cases} 0 & \text{for } x < 1 \\ \frac{1}{3} & \text{for } 1 \leq x < 4 \\ \frac{1}{2} & \text{for } 4 \leq x < 6 \\ \frac{5}{6} & \text{for } 6 \leq x < 10 \\ 1 & \text{for } x \geq 10 \end{cases}$$

find

(a) $P(2 < X \leq 6)$;

(b) $P(X = 4)$;

(c) the probability distribution of X .

12. Find the distribution function of the random variable that has the probability distribution

$$f(x) = \frac{x}{15} \text{ for } x = 1, 2, 3, 4, 5$$

13. If X has the distribution function

$$F(x) = \begin{cases} 0 & \text{for } x < -1 \\ \frac{1}{4} & \text{for } -1 \leq x < 1 \\ \frac{1}{2} & \text{for } 1 \leq x < 3 \\ \frac{3}{4} & \text{for } 3 \leq x < 5 \\ 1 & \text{for } x \geq 5 \end{cases}$$

find

- (a) $P(X \leq 3)$; (b) $P(X = 3)$; (c) $P(X < 3)$;
 (d) $P(X \geq 1)$; (e) $P(-0.4 < X < 4)$; (f) $P(X = 5)$.

14. With reference to Example 4, verify that the values of the distribution function are given by

$$F(x) = \frac{x^2 + 5x}{50}$$

for $x = 1, 2, 3, 4$, and 5 .

15. With reference to Theorem 3, verify that

- (a) $P(X > x_i) = 1 - F(x_i)$ for $i = 1, 2, 3, \dots, n$;
 (b) $P(X \geq x_i) = 1 - F(x_{i-1})$ for $i = 2, 3, \dots, n$,
 and $P(X \geq x_1) = 1$.

3 Continuous Random Variables

In Section 1 we introduced the concept of a random variable as a real-valued function defined over the points of a sample space with a probability measure, and in Figure 1 we illustrated this by assigning the total rolled with a pair of dice to each of the 36 equally likely points of the sample space. In the continuous case, where random variables can take on values on a continuous scale, the procedure is very much the same. The outcomes of experiments are represented by the points on line segments or lines, and the values of random variables are numbers appropriately assigned to the points by means of rules or equations. When the value of a random variable is given directly by a measurement or observation, we generally do not bother to distinguish between the value of the random variable (the measurement that we obtain) and the outcome of the experiment (the corresponding point on the real axis). Thus, if an experiment consists of determining the actual content of a 230-gram jar of instant coffee, the result itself, say, 225.3 grams, is the value of the random variable with which we are concerned, and there is no real need to add that the sample space consists of a certain continuous interval of points on the positive real axis.

The problem of defining probabilities in connection with continuous sample spaces and continuous random variables involves some complications. To illustrate, let us consider the following situation.

EXAMPLE 8

Suppose that we are concerned with the possibility that an accident will occur on a freeway that is 200 kilometers long and that we are interested in the probability that it will occur at a given location, or perhaps on a given stretch of the road. The sample space of this “experiment” consists of a continuum of points, those on the interval from 0 to 200, and we shall assume, for the sake of argument, that the probability that an accident will occur on any interval of length d is $\frac{d}{200}$, with d measured in

kilometers. Note that this assignment of probabilities is consistent with Postulates 1 and 2. (Postulate 1 states that probability of an event is a nonnegative real number; that is, $P(A) \geq 0$ for any subset A of S but in Postulate 2 $P(S) = 1$.) The probabilities $\frac{d}{200}$ are all nonnegative and $P(S) = \frac{200}{200} = 1$. So far this assignment of probabilities applies only to intervals on the line segment from 0 to 200, but if we use Postulate 3 (Postulate 3: If A_1, A_2, A_3, \dots , is a finite or infinite sequence of mutually exclusive events of S , then $P(A_1 \cup A_2 \cup A_3 \cup \dots) = P(A_1) + P(A_2) + P(A_3) + \dots$), we can also obtain probabilities for the union of any finite or countably infinite sequence of nonoverlapping intervals. For instance, the probability that an accident will occur on either of two nonoverlapping intervals of length d_1 and d_2 is

$$\frac{d_1 + d_2}{200}$$

and the probability that it will occur on any one of a countably infinite sequence of nonoverlapping intervals of length d_1, d_2, d_3, \dots is

$$\frac{d_1 + d_2 + d_3 + \dots}{200}$$

With reference to Example 8, observe also that the probability of the accident occurring on a very short interval, say, an interval of 1 centimeter, is only 0.00000005, which is very small. As the length of the interval approaches zero, the probability that an accident will occur on it also approaches zero; indeed, in the continuous case we always assign zero probability to individual points. This does not mean that the corresponding events cannot occur; after all, when an accident occurs on the 200-kilometer stretch of road, it has to occur at some point even though each point has zero probability.

4 Probability Density Functions

The way in which we assigned probabilities in Example 8 is very special, and it is similar in nature to the way in which we assign equal probabilities to the six faces of a die, heads and tails, the 52 playing cards in a standard deck, and so forth. To treat the problem of associating probabilities with values of continuous random variables more generally, suppose that a bottler of soft drinks is concerned about the actual amount of a soft drink that his bottling machine puts into 16-ounce bottles. Evidently, the amount will vary somewhat from bottle to bottle; it is, in fact, a continuous random variable. However, if he rounds the amounts to the nearest tenth of an ounce, he will be dealing with a discrete random variable that has a probability distribution, and this probability distribution may be pictured as a histogram in which the probabilities are given by the areas of rectangles, say, as in the diagram at the top of Figure 6. If he rounds the amounts to the nearest hundredth of an ounce, he will again be dealing with a discrete random variable (a different one) that has a probability distribution, and this probability distribution may be pictured as a probability histogram in which the probabilities

Probability Distributions and Probability Densities

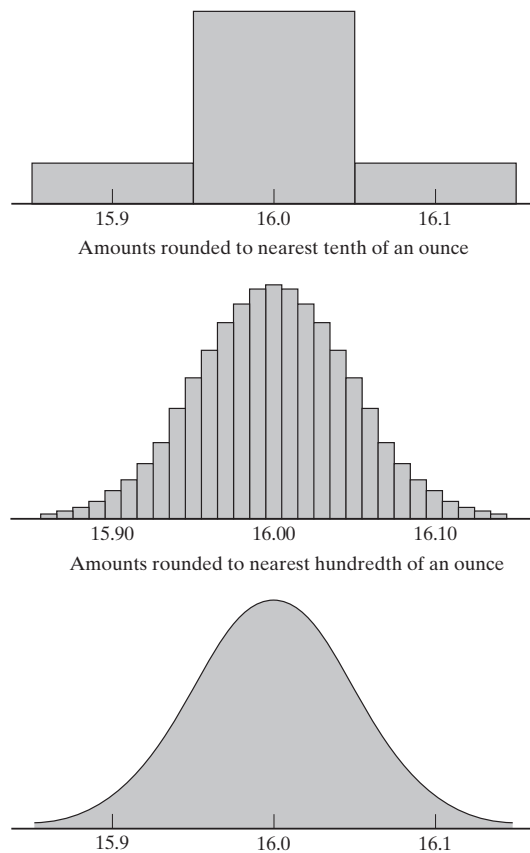


Figure 6. Definition of probability in the continuous case.

are given by the areas of rectangles, say, as in the diagram in the middle of Figure 6.

It should be apparent that if he rounds the amounts to the nearest thousandth of an ounce or to the nearest ten-thousandth of an ounce, the probability histograms of the probability distributions of the corresponding discrete random variables will approach the continuous curve shown in the diagram at the bottom of Figure 6, and the sum of the areas of the rectangles that represent the probability that the amount falls within any specified interval approaches the corresponding area under the curve.

Indeed, the definition of probability in the continuous case presumes for each random variable the existence of a function, called a **probability density function**, such that areas under the curve give the probabilities associated with the corresponding intervals along the horizontal axis. In other words, a probability density function, integrated from a to b (with $a \leq b$), gives the probability that the corresponding random variable will take on a value on the interval from a to b .

DEFINITION 4. PROBABILITY DENSITY FUNCTION. A function with values $f(x)$, defined over the set of all real numbers, is called a **probability density function** of the continuous random variable X if and only if

$$P(a \leq X \leq b) = \int_a^b f(x) dx$$

for any real constants a and b with $a \leq b$.

Probability density functions are also referred to, more briefly, as **probability densities, density functions, densities, or p.d.f.'s**.

Note that $f(c)$, the value of the probability density of X at c , does not give $P(X = c)$ as in the discrete case. In connection with continuous random variables, probabilities are always associated with intervals and $P(X = c) = 0$ for any real constant c . This agrees with what we said on the previous page and it also follows directly from Definition 4 with $a = b = c$.

Because of this property, the value of a probability density function can be changed for some of the values of a random variable without changing the probabilities, and this is why we said in Definition 4 that $f(x)$ is the value of a probability density, not *the* probability density, of the random variable X at x . Also, in view of this property, it does not matter whether we include the endpoints of the interval from a to b ; symbolically, we have the following theorem.

THEOREM 4. If X is a continuous random variable and a and b are real constants with $a \leq b$, then

$$P(a \leq X \leq b) = P(a \leq X < b) = P(a < X \leq b) = P(a < X < b)$$

Analogous to Theorem 1, let us now state the following properties of probability densities, which again follow directly from the postulates of probability.

THEOREM 5. A function can serve as a probability density of a continuous random variable X if its values, $f(x)$, satisfy the conditions[†]

1. $f(x) \geq 0$ for $-\infty < x < \infty$;
2. $\int_{-\infty}^{\infty} f(x) dx = 1$.

EXAMPLE 9

If X has the probability density

$$f(x) = \begin{cases} k \cdot e^{-3x} & \text{for } x > 0 \\ 0 & \text{elsewhere} \end{cases}$$

find k and $P(0.5 \leq X \leq 1)$.

[†]The conditions are not “if and only if” as in Theorem 1 because $f(x)$ could be negative for some values of the random variable without affecting any of the probabilities. However, both conditions of Theorem 5 will be satisfied by nearly all the probability densities used in practice and studied in this text.

Solution

To satisfy the second condition of Theorem 5, we must have

$$\int_{-\infty}^{\infty} f(x) dx = \int_0^{\infty} k \cdot e^{-3x} dx = k \cdot \lim_{t \rightarrow \infty} \left. \frac{e^{-3x}}{-3} \right|_0^t = \frac{k}{3} = 1$$

and it follows that $k = 3$. For the probability we get

$$P(0.5 \leq X \leq 1) = \int_{0.5}^1 3e^{-3x} dx = -e^{-3x} \Big|_{0.5}^1 = -e^{-3} + e^{-1.5} = 0.173$$

Although the random variable of the preceding example cannot take on negative values, we artificially extended the domain of its probability density to include all the real numbers. This is a practice we shall follow throughout this text.

As in the discrete case, there are many problems in which it is of interest to know the probability that the value of a continuous random variable X is less than or equal to some real number x . Thus, let us make the following definition analogous to Definition 3.

DEFINITION 5. DISTRIBUTION FUNCTION. If X is a continuous random variable and the value of its probability density at t is $f(t)$, then the function given by

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt \quad \text{for } -\infty < x < \infty$$

is called the **distribution function** or the **cumulative distribution function** of X .

The properties of distribution functions given in Theorem 2 hold also for the continuous case; that is, $F(-\infty) = 0$, $F(\infty) = 1$, and $F(a) \leq F(b)$ when $a < b$. Furthermore, based on Definition 5, we can state the following theorem.

THEOREM 6. If $f(x)$ and $F(x)$ are the values of the probability density and the distribution function of X at x , then

$$P(a \leq X \leq b) = F(b) - F(a)$$

for any real constants a and b with $a \leq b$, and

$$f(x) = \frac{dF(x)}{dx}$$

where the derivative exists.

EXAMPLE 10

Find the distribution function of the random variable X of Example 9, and use it to reevaluate $P(0.5 \leq X \leq 1)$.

Solution

For $x > 0$,

$$F(x) = \int_{-\infty}^x f(t) dt = \int_0^x 3e^{-3t} dt = -e^{-3t} \Big|_0^x = 1 - e^{-3x}$$

and since $F(x) = 0$ for $x \leq 0$, we can write

$$F(x) = \begin{cases} 0 & \text{for } x \leq 0 \\ 1 - e^{-3x} & \text{for } x > 0 \end{cases}$$

To determine the probability $P(0.5 \leq X \leq 1)$, we use the first part of Theorem 6, getting

$$\begin{aligned} P(0.5 \leq X \leq 1) &= F(1) - F(0.5) \\ &= (1 - e^{-3}) - (1 - e^{-1.5}) \\ &= 0.173 \end{aligned}$$

This agrees with the result obtained by using the probability density directly in Example 9.

EXAMPLE 11

Find a probability density function for the random variable whose distribution function is given by

$$F(x) = \begin{cases} 0 & \text{for } x \leq 0 \\ x & \text{for } 0 < x < 1 \\ 1 & \text{for } x \geq 1 \end{cases}$$

and plot its graph.

Solution

Since the given density function is differentiable everywhere except at $x = 0$ and $x = 1$, we differentiate for $x < 0$, $0 < x < 1$, and $x > 1$, getting 0, 1, and 0. Thus, according to the second part of Theorem 6, we can write

$$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } 0 < x < 1 \\ 0 & \text{for } x > 1 \end{cases}$$

To fill the gaps at $x = 0$ and $x = 1$, we let $f(0)$ and $f(1)$ both equal zero. Actually, it does not matter how the probability density is defined at these two points, but there are certain advantages for choosing the values in such a way that the probability density is nonzero over an open interval. Thus, we can write the probability density of the original random variable as

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

Its graph is shown in Figure 7.

Probability Distributions and Probability Densities

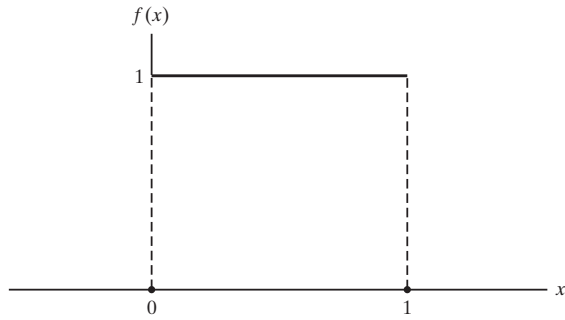


Figure 7. Probability density of Example 11.

In most applications we encounter random variables that are either discrete or continuous, so the corresponding distribution functions have a steplike appearance as in Figure 5, or they are continuous curves or combinations of lines as in Figure 8, which shows the graph of the distribution function of Example 11.

Discontinuous distribution functions like the one shown in Figure 9 arise when random variables are **mixed**. Such a distribution function will be discontinuous at each point having a nonzero probability and continuous elsewhere. As in the discrete case, the height of the step at a point of discontinuity gives the probability that

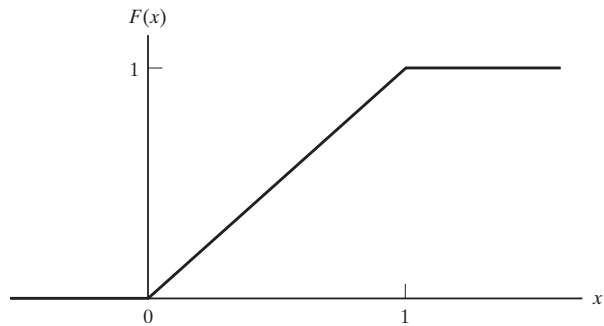


Figure 8. Distribution function of Example 11.

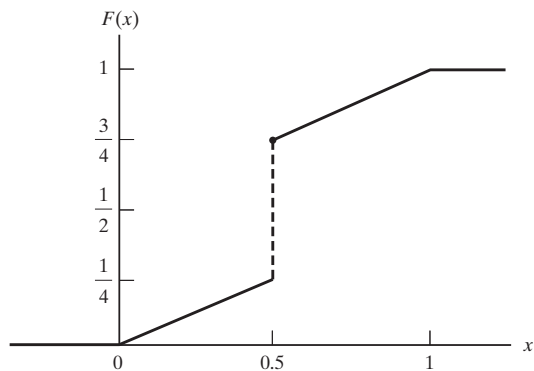


Figure 9. Distribution function of a mixed random variable.

the random variable will take on that particular value. With reference to Figure 9, $P(X = 0.5) = \frac{3}{4} - \frac{1}{4} = \frac{1}{2}$, but otherwise the random variable is like a continuous random variable.

In this chapter we shall limit ourselves to random variables that are discrete or continuous with the latter having distribution functions that are differentiable for all but a finite set of values of the random variables.

Exercises

16. Find the distribution function of the random variable X of Exercise 17 and use it to reevaluate part (b).

17. The probability density of the continuous random variable X is given by

$$f(x) = \begin{cases} \frac{1}{5} & \text{for } 2 < x < 7 \\ 0 & \text{elsewhere} \end{cases}$$

(a) Draw its graph and verify that the total area under the curve (above the x -axis) is equal to 1.

(b) Find $P(3 < X < 5)$.

18. (a) Show that

$$f(x) = e^{-x} \quad \text{for } 0 < x < \infty$$

represents a probability density function.

(b) Sketch a graph of this function and indicate the area associated with the probability that $x > 1$.

(c) Calculate the probability that $x > 1$.

19. (a) Show that

$$f(x) = 3x^2 \quad \text{for } 0 < x < 1$$

represents a density function.

(b) Sketch a graph of this function, and indicate the area associated with the probability that $0.1 < x < 0.5$.

(c) Calculate the probability that $0.1 < x < 0.5$.

20. The probability density of the random variable Y is given by

$$f(y) = \begin{cases} \frac{1}{8}(y+1) & \text{for } 2 < y < 4 \\ 0 & \text{elsewhere} \end{cases}$$

Find $P(Y < 3.2)$ and $P(2.9 < Y < 3.2)$.

21. Find the distribution function of the random variable Y of Exercise 20 and use it to determine the two probabilities asked for in that exercise.

22. The p.d.f. of the random variable X is given by

$$f(x) = \begin{cases} \frac{c}{\sqrt{x}} & \text{for } 0 < x < 4 \\ 0 & \text{elsewhere} \end{cases}$$

Find

(a) the value of c ;

(b) $P(X < \frac{1}{4})$ and $P(X > 1)$.

23. Find the distribution function of the random variable X of Exercise 22 and use it to determine the two probabilities asked for in part (b) of that exercise.

24. The probability density of the random variable Z is given by

$$f(z) = \begin{cases} kze^{-z^2} & \text{for } z > 0 \\ 0 & \text{for } z \leq 0 \end{cases}$$

Find k and draw the graph of this probability density.

25. With reference to Exercise 24, find the distribution function of Z and draw its graph.

26. The density function of the random variable X is given by

$$g(x) = \begin{cases} 6x(1-x) & \text{for } 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$$

Find $P(X < \frac{1}{4})$ and $P(X > \frac{1}{2})$.

27. With reference to Exercise 26, find the distribution function of X and use it to reevaluate the two probabilities asked for in that exercise.

28. Find the distribution function of the random variable X whose probability density is given by

$$f(x) = \begin{cases} x & \text{for } 0 < x < 1 \\ 2-x & \text{for } 1 \leq x < 2 \\ 0 & \text{elsewhere} \end{cases}$$

Probability Distributions and Probability Densities

Also sketch the graphs of the probability density and distribution functions.

29. Find the distribution function of the random variable X whose probability density is given by

$$f(x) = \begin{cases} \frac{1}{3} & \text{for } 0 < x < 1 \\ \frac{1}{3} & \text{for } 2 < x < 4 \\ 0 & \text{elsewhere} \end{cases}$$

Also sketch the graphs of the probability density and distribution functions.

30. With reference to Exercise 28, find $P(0.8 < X < 1.2)$ using

- (a) the probability density;
(b) the distribution function.

31. Find the distribution function of the random variable X whose probability density is given by

$$f(x) = \begin{cases} \frac{x}{2} & \text{for } 0 < x \leq 1 \\ \frac{1}{2} & \text{for } 1 < x \leq 2 \\ \frac{3-x}{2} & \text{for } 2 < x < 3 \\ 0 & \text{elsewhere} \end{cases}$$

Also sketch the graphs of these probability density and distribution functions.

32. The distribution function of the random variable X is given by

$$F(x) = \begin{cases} 0 & \text{for } x < -1 \\ \frac{x+1}{2} & \text{for } -1 \leq x < 1 \\ 1 & \text{for } x \geq 1 \end{cases}$$

Find $P(-\frac{1}{2} < X < \frac{1}{2})$ and $P(2 < X < 3)$.

33. With reference to Exercise 32, find the probability density of X and use it to recalculate the two probabilities.

34. The distribution function of the random variable Y is given by

$$F(y) = \begin{cases} 1 - \frac{9}{y^2} & \text{for } y > 3 \\ 0 & \text{elsewhere} \end{cases}$$

Find $P(Y \leq 5)$ and $P(Y > 8)$.

35. With reference to Exercise 34, find the probability density of Y and use it to recalculate the two probabilities.

36. With reference to Exercise 34 and the result of Exercise 35, sketch the graphs of the distribution function and the probability density of Y , letting $f(3) = 0$.

37. The distribution function of the random variable X is given by

$$F(x) = \begin{cases} 1 - (1+x)e^{-x} & \text{for } x > 0 \\ 0 & \text{for } x \leq 0 \end{cases}$$

Find $P(X \leq 2)$, $P(1 < X < 3)$, and $P(X > 4)$.

38. With reference to Exercise 37, find the probability density of X .

39. With reference to Figure 9, find expressions for the values of the distribution function of the mixed random variable X for

- (a) $x \leq 0$; (b) $0 < x < 0.5$;
(c) $0.5 \leq x < 1$; (d) $x \geq 1$.

40. Use the results of Exercise 39 to find expressions for the values of the probability density of the mixed random variable X for

- (a) $x < 0$; (b) $0 < x < 0.5$;
(c) $0.5 < x < 1$; (d) $x > 1$.

$P(X = 0.5) = \frac{1}{2}$, and $f(0)$ and $f(1)$ are undefined.

41. The distribution function of the mixed random variable Z is given by

$$F(z) = \begin{cases} 0 & \text{for } z < -2 \\ \frac{z+4}{8} & \text{for } -2 \leq z < 2 \\ 1 & \text{for } z \geq 2 \end{cases}$$

Find $P(Z = -2)$, $P(Z = 2)$, $P(-2 < Z < 1)$, and $P(0 \leq Z \leq 2)$.

5 Multivariate Distributions

In the beginning of this chapter we defined a random variable as a real-valued function defined over a sample space with a probability measure, and it stands to reason that many different random variables can be defined over one and the same sample space. With reference to the sample space of Figure 1, for example, we considered only the random variable whose values were the totals rolled with a pair of dice, but we could also have considered the random variable whose values are the products of the numbers rolled with the two dice, the random variable whose values are the differences between the numbers rolled with the red die and the green die, the random variable whose values are 0, 1, or 2 depending on the number of dice that come up 2, and so forth. Closer to life, an experiment may consist of randomly choosing some of the 345 students attending an elementary school, and the principal may be interested in their I.Q.'s, the school nurse in their weights, their teachers in the number of days they have been absent, and so forth.

In this section we shall be concerned first with the **bivariate case**, that is, with situations where we are interested at the same time in a pair of random variables defined over a joint sample space. Later, we shall extend this discussion to the **multivariate case**, covering any finite number of random variables.

If X and Y are discrete random variables, we write the probability that X will take on the value x and Y will take on the value y as $P(X = x, Y = y)$. Thus, $P(X = x, Y = y)$ is the probability of the intersection of the events $X = x$ and $Y = y$. As in the **univariate case**, where we dealt with one random variable and could display the probabilities associated with all values of X by means of a table, we can now, in the bivariate case, display the probabilities associated with all pairs of values of X and Y by means of a table.

EXAMPLE 12

Two caplets are selected at random from a bottle containing 3 aspirin, 2 sedative, and 4 laxative caplets. If X and Y are, respectively, the numbers of aspirin and sedative caplets included among the 2 caplets drawn from the bottle, find the probabilities associated with all possible pairs of values of X and Y .

Solution

The possible pairs are (0, 0), (0, 1), (1, 0), (1, 1), (0, 2), and (2, 0). To find the probability associated with (1, 0), for example, observe that we are concerned with the event of getting one of the 3 aspirin caplets, none of the 2 sedative caplets, and hence, one of the 4 laxative caplets. The number of ways in which this can be done is $\binom{3}{1} \binom{2}{0} \binom{4}{1} = 12$, and the total number of ways in which 2 of the 9 caplets can be selected is $\binom{9}{2} = 36$. Since those possibilities are all equally likely by virtue of the assumption that the selection is random, it follows from a theorem (If an experiment can result in any one of N different equally likely outcomes, and if n of these outcomes together constitute event A , then the probability of event A is $P(A) = n/N$) that the probability associated with (1, 0) is $\frac{12}{36} = \frac{1}{3}$. Similarly, the probability associated with (1, 1) is

$$\frac{\binom{3}{1} \binom{2}{1} \binom{4}{0}}{36} = \frac{6}{36} = \frac{1}{6}$$

and, continuing this way, we obtain the values shown in the following table:

		x		
		0	1	2
y	0	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{12}$
	1	$\frac{2}{9}$	$\frac{1}{6}$	
	2	$\frac{1}{36}$		

Actually, as in the univariate case, it is generally preferable to represent probabilities such as these by means of a formula. In other words, it is preferable to express the probabilities by means of a function with the values $f(x, y) = P(X = x, Y = y)$ for any pair of values (x, y) within the range of the random variables X and Y . For instance, for the two random variables of Example 12 we can write

$$f(x, y) = \frac{\binom{3}{x} \binom{2}{y} \binom{4}{2-x-y}}{\binom{9}{2}} \quad \text{for } x = 0, 1, 2; \quad y = 0, 1, 2; \\ 0 \leq x + y \leq 2$$

DEFINITION 6. JOINT PROBABILITY DISTRIBUTION. If X and Y are discrete random variables, the function given by $f(x, y) = P(X = x, Y = y)$ for each pair of values (x, y) within the range of X and Y is called the **joint probability distribution** of X and Y .

Analogous to Theorem 1, let us state the following theorem, which follows from the postulates of probability.

THEOREM 7. A bivariate function can serve as the joint probability distribution of a pair of discrete random variables X and Y if and only if its values, $f(x, y)$, satisfy the conditions

1. $f(x, y) \geq 0$ for each pair of values (x, y) within its domain;
2. $\sum_x \sum_y f(x, y) = 1$, where the double summation extends over all possible pairs (x, y) within its domain.

EXAMPLE 13

Determine the value of k for which the function given by

$$f(x, y) = kxy \quad \text{for } x = 1, 2, 3; \quad y = 1, 2, 3$$

can serve as a joint probability distribution.

DEFINITION 8. JOINT PROBABILITY DENSITY FUNCTION. A bivariate function with values $f(x, y)$ defined over the xy -plane is called a **joint probability density function** of the continuous random variables X and Y if and only if

$$P(X, Y) \in A = \iint_A f(x, y) dx dy$$

for any region A in the xy -plane.

Analogous to Theorem 5, it follows from the postulates of probability that

THEOREM 8. A bivariate function can serve as a joint probability density function of a pair of continuous random variables X and Y if its values, $f(x, y)$, satisfy the conditions

1. $f(x, y) \geq 0$ for $-\infty < x < \infty$, $-\infty < y < \infty$;
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1$.

EXAMPLE 15

Given the joint probability density function

$$f(x, y) = \begin{cases} \frac{3}{5}x(y+x) & \text{for } 0 < x < 1, 0 < y < 2 \\ 0 & \text{elsewhere} \end{cases}$$

of two random variables X and Y , find $P[(X, Y) \in A]$, where A is the region $\{(x, y) | 0 < x < \frac{1}{2}, 1 < y < 2\}$.

Solution

$$\begin{aligned} P[(X, Y) \in A] &= P\left(0 < X < \frac{1}{2}, 1 < Y < 2\right) \\ &= \int_1^2 \int_0^{\frac{1}{2}} \frac{3}{5}x(y+x) dx dy \\ &= \int_1^2 \left. \frac{3x^2y}{10} + \frac{3x^3}{15} \right|_{x=0}^{x=\frac{1}{2}} dy \\ &= \int_1^2 \left(\frac{3y}{40} + \frac{1}{40}\right) dy = \left. \frac{3y^2}{80} + \frac{y}{40} \right|_1^2 \\ &= \frac{11}{80} \end{aligned}$$

Analogous to Definition 7, we have the following definition of the joint distribution function of two continuous random variables.

DEFINITION 9. JOINT DISTRIBUTION FUNCTION. *If X and Y are continuous random variables, the function given by*

$$F(x, y) = P(X \leq x, Y \leq y) = \int_{-\infty}^y \int_{-\infty}^x f(s, t) \, ds \, dt \quad \begin{array}{l} \text{for } -\infty < x < \infty, \\ -\infty < y < \infty \end{array}$$

where $f(s, t)$ is the joint probability density of X and Y at (s, t) , is called the **joint distribution function of X and Y** .

Note that the properties of joint distribution functions, which the reader will be asked to prove in Exercise 48 for the discrete case, hold also for the continuous case.

As in Section 4, we shall limit our discussion here to random variables whose joint distribution function is continuous everywhere and partially differentiable with respect to each variable for all but a finite set of values of the two random variables.

Analogous to the relationship $f(x) = \frac{dF(x)}{dx}$ of Theorem 6, partial differentiation in Definition 9 leads to

$$f(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y)$$

wherever these partial derivatives exist. As in Section 4, the joint distribution function of two continuous random variables determines their **joint density** (short for joint probability density function) at all points (x, y) where the joint density is continuous. Also as in Section 4, we generally let the values of joint probability densities equal zero wherever they are not defined by the above relationship.

EXAMPLE 16

If the joint probability density of X and Y is given by

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

find the joint distribution function of these two random variables.

Solution

If either $x < 0$ or $y < 0$, it follows immediately that $F(x, y) = 0$. For $0 < x < 1$ and $0 < y < 1$ (Region I of Figure 10), we get

$$F(x, y) = \int_0^y \int_0^x (s + t) \, ds \, dt = \frac{1}{2}xy(x + y)$$

for $x > 1$ and $0 < y < 1$ (Region II of Figure 10), we get

$$F(x, y) = \int_0^y \int_0^1 (s + t) \, ds \, dt = \frac{1}{2}y(y + 1)$$

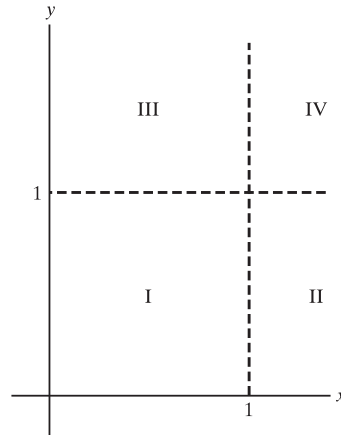


Figure 10. Diagram for Example 16.

for $0 < x < 1$ and $y > 1$ (Region III of Figure 10), we get

$$F(x, y) = \int_0^1 \int_0^x (s+t) ds dt = \frac{1}{2}x(x+1)$$

and for $x > 1$ and $y > 1$ (Region IV of Figure 10), we get

$$F(x, y) = \int_0^1 \int_0^1 (s+t) ds dt = 1$$

Since the joint distribution function is everywhere continuous, the boundaries between any two of these regions can be included in either one, and we can write

$$F(x, y) = \begin{cases} 0 & \text{for } x \leq 0 \text{ or } y \leq 0 \\ \frac{1}{2}xy(x+y) & \text{for } 0 < x < 1, 0 < y < 1 \\ \frac{1}{2}y(y+1) & \text{for } x \geq 1, 0 < y < 1 \\ \frac{1}{2}x(x+1) & \text{for } 0 < x < 1, y \geq 1 \\ 1 & \text{for } x \geq 1, y \geq 1 \end{cases}$$

EXAMPLE 17

Find the joint probability density of the two random variables X and Y whose joint distribution function is given by

$$F(x, y) = \begin{cases} (1 - e^{-x})(1 - e^{-y}) & \text{for } x > 0 \text{ and } y > 0 \\ 0 & \text{elsewhere} \end{cases}$$

Also use the joint probability density to determine $P(1 < X < 3, 1 < Y < 2)$.

Solution

Since partial differentiation yields

$$\frac{\partial^2}{\partial x \partial y} F(x, y) = e^{-(x+y)}$$

for $x > 0$ and $y > 0$ and 0 elsewhere, we find that the joint probability density of X and Y is given by

$$f(x, y) = \begin{cases} e^{-(x+y)} & \text{for } x > 0 \text{ and } y > 0 \\ 0 & \text{elsewhere} \end{cases}$$

Thus, integration yields

$$\begin{aligned} \int_1^2 \int_1^3 e^{-(x+y)} dx dy &= (e^{-1} - e^{-3})(e^{-1} - e^{-2}) \\ &= e^{-2} - e^{-3} - e^{-4} + e^{-5} \\ &= 0.074 \end{aligned}$$

for $P(1 < X < 3, 1 < Y < 2)$.

For two random variables, the joint probability is, geometrically speaking, a surface, and the probability that we calculated in the preceding example is given by the volume under this surface, as shown in Figure 11.

All the definitions of this section can be generalized to the **multivariate** case, where there are n random variables. Corresponding to Definition 6, the values of the joint probability distribution of n discrete random variables X_1, X_2, \dots, X_n are given by

$$f(x_1, x_2, \dots, x_n) = P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$$

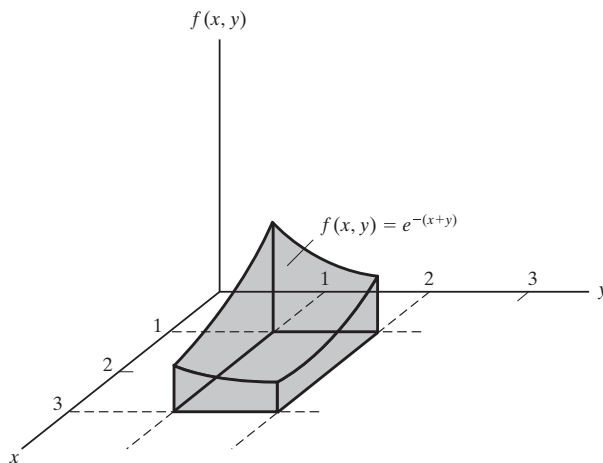


Figure 11. Diagram for Example 17.

Probability Distributions and Probability Densities

for each n -tuple (x_1, x_2, \dots, x_n) within the range of the random variables; and corresponding to Definition 7, the values of their joint distribution function are given by

$$F(x_1, x_2, \dots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n)$$

for $-\infty < x_1 < \infty, -\infty < x_2 < \infty, \dots, -\infty < x_n < \infty$.

EXAMPLE 18

If the joint probability distribution of three discrete random variables $X, Y,$ and Z is given by

$$f(x, y, z) = \frac{(x+y)z}{63} \quad \text{for } x = 1, 2; \quad y = 1, 2, 3; \quad z = 1, 2$$

find $P(X = 2, Y + Z \leq 3)$.

Solution

$$\begin{aligned} P(X = 2, Y + Z \leq 3) &= f(2, 1, 1) + f(2, 1, 2) + f(2, 2, 1) \\ &= \frac{3}{63} + \frac{6}{63} + \frac{4}{63} \\ &= \frac{13}{63} \end{aligned}$$

In the continuous case, probabilities are again obtained by integrating the joint probability density, and the joint distribution function is given by

$$F(x_1, x_2, \dots, x_n) = \int_{-\infty}^{x_n} \dots \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} f(t_1, t_2, \dots, t_n) dt_1 dt_2 \dots dt_n$$

for $-\infty < x_1 < \infty, -\infty < x_2 < \infty, \dots, -\infty < x_n < \infty$, analogous to Definition 9. Also, partial differentiation yields

$$f(x_1, x_2, \dots, x_n) = \frac{\partial^n}{\partial x_1 \partial x_2 \dots \partial x_n} F(x_1, x_2, \dots, x_n)$$

wherever these partial derivatives exist.

EXAMPLE 19

If the **trivariate** probability density of $X_1, X_2,$ and X_3 is given by

$$f(x_1, x_2, x_3) = \begin{cases} (x_1 + x_2)e^{-x_3} & \text{for } 0 < x_1 < 1, 0 < x_2 < 1, x_3 > 0 \\ 0 & \text{elsewhere} \end{cases}$$

find $P[(X_1, X_2, X_3) \in A]$, where A is the region

$$\left\{ (x_1, x_2, x_3) \mid 0 < x_1 < \frac{1}{2}, \frac{1}{2} < x_2 < 1, x_3 < 1 \right\}$$

Solution

$$\begin{aligned}
 P[(X_1, X_2, X_3) \in A] &= P\left(0 < X_1 < \frac{1}{2}, \frac{1}{2} < X_2 < 1, X_3 < 1\right) \\
 &= \int_0^1 \int_{\frac{1}{2}}^1 \int_0^{\frac{1}{2}} (x_1 + x_2)e^{-x_3} dx_1 dx_2 dx_3 \\
 &= \int_0^1 \int_{\frac{1}{2}}^1 \left(\frac{1}{8} + \frac{x_2}{2}\right) e^{-x_3} dx_2 dx_3 \\
 &= \int_0^1 \frac{1}{4} e^{-x_3} dx_3 \\
 &= \frac{1}{4}(1 - e^{-1}) = 0.158
 \end{aligned}$$

Exercises

42. If the values of the joint probability distribution of X and Y are as shown in the table

		x		
		0	1	2
y	0	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{24}$
	1	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{40}$
	2	$\frac{1}{8}$	$\frac{1}{20}$	
	3	$\frac{1}{120}$		

find

- (a) $P(X = 1, Y = 2)$; (b) $P(X = 0, 1 \leq Y < 3)$;
 (c) $P(X + Y \leq 1)$; (d) $P(X > Y)$.

43. With reference to Exercise 42, find the following values of the joint distribution function of the two random variables:

- (a) $F(1.2, 0.9)$; (b) $F(-3, 1.5)$;
 (c) $F(2, 0)$; (d) $F(4, 2.7)$.

44. If the joint probability distribution of X and Y is given by

$$f(x, y) = c(x^2 + y^2) \quad \text{for } x = -1, 0, 1, 3; \quad y = -1, 2, 3$$

find the value of c .

45. With reference to Exercise 44 and the value obtained for c , find

- (a) $P(X \leq 1, Y > 2)$;
 (b) $P(X = 0, Y \leq 2)$;
 (c) $P(X + Y > 2)$.

46. Show that there is no value of k for which

$$f(x, y) = ky(2y - x) \quad \text{for } x = 0, 3; \quad y = 0, 1, 2$$

can serve as the joint probability distribution of two random variables.

47. If the joint probability distribution of X and Y is given by

$$f(x, y) = \frac{1}{30}(x + y) \quad \text{for } x = 0, 1, 2, 3; \quad y = 0, 1, 2$$

construct a table showing the values of the joint distribution function of the two random variables at the 12 points $(0, 0), (0, 1), \dots, (3, 2)$.

48. If $F(x, y)$ is the value of the joint distribution function of two discrete random variables X and Y at (x, y) , show that

- (a) $F(-\infty, -\infty) = 0$;
 (b) $F(\infty, \infty) = 1$;
 (c) if $a < b$ and $c < d$, then $F(a, c) \leq F(b, d)$.

49. Determine k so that

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

can serve as a joint probability density.

50. If the joint probability density of X and Y is given by

$$f(x, y) = \begin{cases} 24xy & \text{for } 0 < x < 1, 0 < y < 1, x + y < 1 \\ 0 & \text{elsewhere} \end{cases}$$

find $P(X + Y < \frac{1}{2})$.

51. If the joint probability density of X and Y is given by

$$f(x, y) = \begin{cases} 2 & \text{for } x > 0, y > 0, x + y < 1 \\ 0 & \text{elsewhere} \end{cases}$$

find

- (a) $P(X \leq \frac{1}{2}, Y \leq \frac{1}{2})$;
 (b) $P(X + Y > \frac{2}{3})$;
 (c) $P(X > 2Y)$.

52. With reference to Exercise 51, find an expression for the values of the joint distribution function of X and Y when $x > 0, y > 0$, and $x + y < 1$, and use it to verify the result of part (a).

53. If the joint probability density of X and Y is given by

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

find the probability that the sum of the values of X and Y will exceed $\frac{1}{2}$.

54. Find the joint probability density of the two random variables X and Y whose joint distribution function is given by

$$F(x, y) = \begin{cases} (1 - e^{-x^2})(1 - e^{-y^2}) & \text{for } x > 0, y > 0 \\ 0 & \text{elsewhere} \end{cases}$$

55. Use the joint probability density obtained in Exercise 54 to find $P(1 < X \leq 2, 1 < Y \leq 2)$.

56. Find the joint probability density of the two random variables X and Y whose joint distribution function is given by

$$F(x, y) = \begin{cases} 1 - e^{-x} - e^{-y} + e^{-x-y} & \text{for } x > 0, y > 0 \\ 0 & \text{elsewhere} \end{cases}$$

57. Use the joint probability density obtained in Exercise 56 to find $P(X + Y > 3)$.

58. If $F(x, y)$ is the value of the joint distribution function of the two continuous random variables X and Y at (x, y) , express $P(a < X \leq b, c < Y \leq d)$ in terms of $F(a, c)$, $F(a, d)$, $F(b, c)$, and $F(b, d)$. Observe that the result holds also for discrete random variables.

59. Use the formula obtained in Exercise 58 to verify the result, 0.074, of Example 17.

60. Use the formula obtained in Exercise 58 to verify the result of Exercise 55.

61. Use the formula obtained in Exercise 58 to verify the result of Exercise 57.

62. Find k if the joint probability distribution of X, Y , and Z is given by

$$f(x, y, z) = kxyz$$

for $x = 1, 2; y = 1, 2, 3; z = 1, 2$.

63. With reference to Exercise 62, find

- (a) $P(X = 1, Y \leq 2, Z = 1)$;
 (b) $P(X = 2, Y + Z = 4)$.

64. With reference to Exercise 62, find the following values of the joint distribution function of the three random variables:

- (a) $F(2, 1, 2)$;
 (b) $F(1, 0, 1)$;
 (c) $F(4, 4, 4)$.

65. Find k if the joint probability density of X, Y , and Z is given by

$$f(x, y, z) = \begin{cases} kxy(1-z) & \text{for } 0 < x < 1, 0 < y < 1, \\ & 0 < z < 1, x + y + z < 1 \\ 0 & \text{elsewhere} \end{cases}$$

66. With reference to Exercise 65, find $P(X + Y < \frac{1}{2})$.

67. Use the result of Example 16 to verify that the joint distribution function of the random variables X_1, X_2 , and X_3 of Example 19 is given by

$$F(x_1, x_2, x_3) = \begin{cases} 0 & \text{for } x_1 \leq 0, x_2 \leq 0, \text{ or } x_3 \leq 0 \\ \frac{1}{2}x_1x_2(x_1 + x_2)(1 - e^{-x_3}) & \text{for } 0 < x_1 < 1, 0 < x_2 < 1, x_3 > 0 \\ \frac{1}{2}x_2(x_2 + 1)(1 - e^{-x_3}) & \text{for } x_1 \geq 1, 0 < x_2 < 1, x_3 > 0 \\ \frac{1}{2}x_1(x_1 + 1)(1 - e^{-x_3}) & \text{for } 0 < x_1 < 1, x_2 \geq 1, x_3 > 0 \\ 1 - e^{-x_3} & \text{for } x_1 \geq 1, x_2 \geq 1, x_3 > 0 \end{cases}$$

68. If the joint probability density of X, Y , and Z is given by

$$f(x, y, z) = \begin{cases} \frac{1}{3}(2x + 3y + z) & \text{for } 0 < x < 1, 0 < y < 1, \\ & 0 < z < 1 \\ 0 & \text{elsewhere} \end{cases}$$

find

- (a) $P(X = \frac{1}{2}, Y = \frac{1}{2}, Z = \frac{1}{2})$;
 (b) $P(X < \frac{1}{2}, Y < \frac{1}{2}, Z < \frac{1}{2})$.

6 Marginal Distributions

To introduce the concept of a **marginal distribution**, let us consider the following example.

EXAMPLE 20

In Example 12 we derived the joint probability distribution of two random variables X and Y , the number of aspirin caplets and the number of sedative caplets included among two caplets drawn at random from a bottle containing three aspirin, two sedative, and four laxative caplets. Find the probability distribution of X alone and that of Y alone.

Solution

The results of Example 12 are shown in the following table, together with the marginal totals, that is, the totals of the respective rows and columns:

		x			
		0	1	2	
y	0	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{12}$	$\frac{7}{12}$
	1	$\frac{2}{9}$	$\frac{1}{6}$		$\frac{7}{18}$
	2	$\frac{1}{36}$			$\frac{1}{36}$
		$\frac{5}{12}$	$\frac{1}{2}$	$\frac{1}{12}$	

The column totals are the probabilities that X will take on the values 0, 1, and 2. In other words, they are the values

$$g(x) = \sum_{y=0}^2 f(x, y) \quad \text{for } x = 0, 1, 2$$

of the probability distribution of X . By the same token, the row totals are the values

$$h(y) = \sum_{x=0}^2 f(x, y) \quad \text{for } y = 0, 1, 2$$

of the probability distribution of Y .

We are thus led to the following definition.

DEFINITION 10. MARGINAL DISTRIBUTION. *If X and Y are discrete random variables and $f(x, y)$ is the value of their joint probability distribution at (x, y) , the function given by*

$$g(x) = \sum_y f(x, y)$$

for each x within the range of X is called the **marginal distribution of X** . Correspondingly, the function given by

$$h(y) = \sum_x f(x, y)$$

for each y within the range of Y is called the **marginal distribution of Y** .

When X and Y are continuous random variables, the probability distributions are replaced by probability densities, the summations are replaced by integrals, and we obtain the following definition.

DEFINITION 11. MARGINAL DENSITY. If X and Y are continuous random variables and $f(x, y)$ is the value of their joint probability density at (x, y) , the function given by

$$g(x) = \int_{-\infty}^{\infty} f(x, y) dy \quad \text{for } -\infty < x < \infty$$

is called the **marginal density of X** . Correspondingly, the function given by

$$h(y) = \int_{-\infty}^{\infty} f(x, y) dx \quad \text{for } -\infty < y < \infty$$

is called the **marginal density of Y** .

EXAMPLE 21

Given the joint probability density

$$f(x, y) = \begin{cases} \frac{2}{3}(x + 2y) & \text{for } 0 < x < 1, 0 < y < 1 \\ 0 & \text{elsewhere} \end{cases}$$

find the marginal densities of X and Y .

Solution

Performing the necessary integrations, we get

$$g(x) = \int_{-\infty}^{\infty} f(x, y) dy = \int_0^1 \frac{2}{3}(x + 2y) dy = \frac{2}{3}(x + 1)$$

for $0 < x < 1$ and $g(x) = 0$ elsewhere. Likewise,

$$h(y) = \int_{-\infty}^{\infty} f(x, y) dx = \int_0^1 \frac{2}{3}(x + 2y) dx = \frac{1}{3}(1 + 4y)$$

for $0 < y < 1$ and $h(y) = 0$ elsewhere.

When we are dealing with more than two random variables, we can speak not only of the marginal distributions of the individual random variables, but also of the

joint marginal distributions of several of the random variables. If the joint probability distribution of the discrete random variables X_1, X_2, \dots , and X_n has the values $f(x_1, x_2, \dots, x_n)$, the marginal distribution of X_1 alone is given by

$$g(x_1) = \sum_{x_2} \cdots \sum_{x_n} f(x_1, x_2, \dots, x_n)$$

for all values within the range of X_1 , the joint marginal distribution of X_1, X_2 , and X_3 is given by

$$m(x_1, x_2, x_3) = \sum_{x_4} \cdots \sum_{x_n} f(x_1, x_2, \dots, x_n)$$

for all values within the range of X_1, X_2 , and X_3 , and other marginal distributions can be defined in the same way. For the continuous case, probability distributions are replaced by probability densities, summations are replaced by integrals, and if the joint probability density of the continuous random variables X_1, X_2, \dots , and X_n has the values $f(x_1, x_2, \dots, x_n)$, the marginal density of X_2 alone is given by

$$h(x_2) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, x_2, \dots, x_n) dx_1 dx_3 \cdots dx_n$$

for $-\infty < x_2 < \infty$, the joint marginal density of X_1 and X_n is given by

$$\varphi(x_1, x_n) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, x_2, \dots, x_n) dx_2 dx_3 \cdots dx_{n-1}$$

for $-\infty < x_1 < \infty$ and $-\infty < x_n < \infty$, and so forth.

EXAMPLE 22

Considering again the trivariate probability density of Example 19,

$$f(x_1, x_2, x_3) = \begin{cases} (x_1 + x_2)e^{-x_3} & \text{for } 0 < x_1 < 1, 0 < x_2 < 1, x_3 > 0 \\ 0 & \text{elsewhere} \end{cases}$$

find the joint marginal density of X_1 and X_3 and the marginal density of X_1 alone.

Solution

Performing the necessary integration, we find that the joint marginal density of X_1 and X_3 is given by

$$m(x_1, x_3) = \int_0^1 (x_1 + x_2)e^{-x_3} dx_2 = \left(x_1 + \frac{1}{2}\right) e^{-x_3}$$

for $0 < x_1 < 1$ and $x_3 > 0$ and $m(x_1, x_3) = 0$ elsewhere. Using this result, we find that the marginal density of X_1 alone is given by

$$\begin{aligned} g(x_1) &= \int_0^{\infty} \int_0^1 f(x_1, x_2, x_3) dx_2 dx_3 = \int_0^{\infty} m(x_1, x_3) dx_3 \\ &= \int_0^{\infty} \left(x_1 + \frac{1}{2}\right) e^{-x_3} dx_3 = x_1 + \frac{1}{2} \end{aligned}$$

for $0 < x_1 < 1$ and $g(x_1) = 0$ elsewhere.

Corresponding to the various marginal and joint marginal distributions and densities we have introduced in this section, we can also define **marginal** and **joint marginal distribution functions**. Some problems relating to such distribution functions will be left to the reader in Exercises 72, 79, and 80.

7 Conditional Distributions

In the conditional probability of event A , given event B , as

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

provided $P(B) \neq 0$. Suppose now that A and B are the events $X = x$ and $Y = y$ so that we can write

$$\begin{aligned} P(X = x|Y = y) &= \frac{P(X = x, Y = y)}{P(Y = y)} \\ &= \frac{f(x, y)}{h(y)} \end{aligned}$$

provided $P(Y = y) = h(y) \neq 0$, where $f(x, y)$ is the value of the joint probability distribution of X and Y at (x, y) , and $h(y)$ is the value of the marginal distribution of Y at y . Denoting the conditional probability by $f(x|y)$ to indicate that x is a variable and y is fixed, let us now make the following definition.

DEFINITION 12. CONDITIONAL DISTRIBUTION. If $f(x, y)$ is the value of the joint probability distribution of the discrete random variables X and Y at (x, y) and $h(y)$ is the value of the marginal distribution of Y at y , the function given by

$$f(x|y) = \frac{f(x, y)}{h(y)} \quad h(y) \neq 0$$

for each x within the range of X is called the **conditional distribution of X given $Y = y$** . Correspondingly, if $g(x)$ is the value of the marginal distribution of X at x , the function given by

$$w(y|x) = \frac{f(x, y)}{g(x)} \quad g(x) \neq 0$$

for each y within the range of Y is called the **conditional distribution of Y given $X = x$** .

EXAMPLE 23

With reference to Examples 12 and 20, find the conditional distribution of X given $Y = 1$.

Solution

Substituting the appropriate values from the table in Example 20, we get

$$f(0|1) = \frac{\frac{2}{9}}{\frac{7}{9}} = \frac{2}{7}$$

$$f(1|1) = \frac{\frac{1}{6}}{\frac{7}{18}} = \frac{3}{7}$$

$$f(2|1) = \frac{0}{\frac{7}{18}} = 0$$

When X and Y are continuous random variables, the probability distributions are replaced by probability densities, and we obtain the following definition.

DEFINITION 13. CONDITIONAL DENSITY. If $f(x, y)$ is the value of the joint density of the continuous random variables X and Y at (x, y) and $h(y)$ is the value of the marginal distribution of Y at y , the function given by

$$f(x|y) = \frac{f(x, y)}{h(y)} \quad h(y) \neq 0$$

for $-\infty < x < \infty$, is called the **conditional density of X given $Y = y$** . Correspondingly, if $g(x)$ is the value of the marginal density of X at x , the function given by

$$w(y|x) = \frac{f(x, y)}{g(x)} \quad g(x) \neq 0$$

for $-\infty < y < \infty$, is called the **conditional density of Y given $X = x$** .

EXAMPLE 24

With reference to Example 21, find the conditional density of X given $Y = y$, and use it to evaluate $P(X \leq \frac{1}{2} | Y = \frac{1}{2})$.

Solution

Using the results obtained on the previous page, we have

$$f(x|y) = \frac{f(x, y)}{h(y)} = \frac{\frac{2}{3}(x+2y)}{\frac{1}{3}(1+4y)}$$

$$= \frac{2x+4y}{1+4y}$$

for $0 < x < 1$ and $f(x|y) = 0$ elsewhere. Now,

$$f\left(x \left| \frac{1}{2} \right.\right) = \frac{2x+4 \cdot \frac{1}{2}}{1+4 \cdot \frac{1}{2}}$$

$$= \frac{2x+2}{3}$$

and we can write

$$P\left(X \leq \frac{1}{2} \mid Y = \frac{1}{2}\right) = \int_0^{\frac{1}{2}} \frac{2x+2}{3} dx = \frac{5}{12}$$

It is of interest to note that in Figure 12 this probability is given by the ratio of the area of trapezoid $ABCD$ to the area of trapezoid $AEFD$.

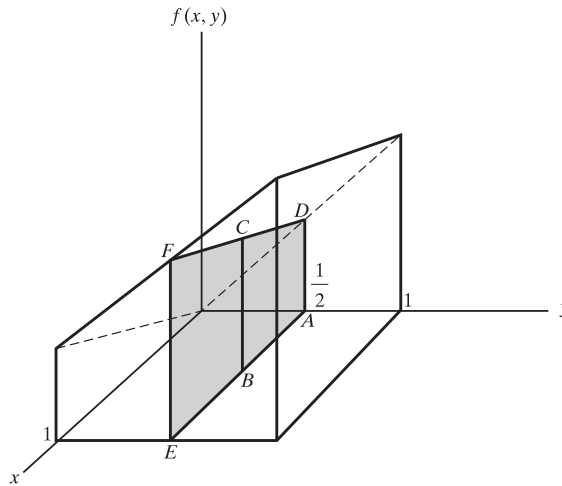


Figure 12. Diagram for Example 24.

EXAMPLE 25

Given the joint probability density

$$f(x, y) = \begin{cases} 4xy & \text{for } 0 < x < 1, 0 < y < 1 \\ 0 & \text{elsewhere} \end{cases}$$

find the marginal densities of X and Y and the conditional density of X given $Y = y$.

Solution

Performing the necessary integrations, we get

$$\begin{aligned} g(x) &= \int_{-\infty}^{\infty} f(x, y) dy = \int_0^1 4xy dy \\ &= 2xy^2 \Big|_{y=0}^{y=1} = 2x \end{aligned}$$

for $0 < x < 1$, and $g(x) = 0$ elsewhere; also

$$\begin{aligned} h(y) &= \int_{-\infty}^{\infty} f(x, y) dx = \int_0^1 4xy dx \\ &= 2x^2y \Big|_{x=0}^{x=1} = 2y \end{aligned}$$

for $0 < y < 1$, and $h(y) = 0$ elsewhere. Then, substituting into the formula for a conditional density, we get

$$f(x|y) = \frac{f(x,y)}{h(y)} = \frac{4xy}{2y} = 2x$$

for $0 < x < 1$, and $f(x|y) = 0$ elsewhere.

When we are dealing with more than two random variables, whether continuous or discrete, we can consider various different kinds of conditional distributions or densities. For instance, if $f(x_1, x_2, x_3, x_4)$ is the value of the joint distribution of the discrete random variables X_1, X_2, X_3 , and X_4 at (x_1, x_2, x_3, x_4) , we can write

$$p(x_3|x_1, x_2, x_4) = \frac{f(x_1, x_2, x_3, x_4)}{g(x_1, x_2, x_4)} \quad g(x_1, x_2, x_4) \neq 0$$

for the value of the conditional distribution of X_3 at x_3 given $X_1 = x_1, X_2 = x_2$, and $X_4 = x_4$, where $g(x_1, x_2, x_4)$ is the value of the joint marginal distribution of X_1, X_2 , and X_4 at (x_1, x_2, x_4) . We can also write

$$q(x_2, x_4|x_1, x_3) = \frac{f(x_1, x_2, x_3, x_4)}{m(x_1, x_3)} \quad m(x_1, x_3) \neq 0$$

for the value of the **joint conditional distribution** of X_2 and X_4 at (x_2, x_4) given $X_1 = x_1$ and $X_3 = x_3$, or

$$r(x_2, x_3, x_4|x_1) = \frac{f(x_1, x_2, x_3, x_4)}{b(x_1)} \quad b(x_1) \neq 0$$

for the value of the joint conditional distribution of X_2, X_3 , and X_4 at (x_2, x_3, x_4) given $X_1 = x_1$.

When we are dealing with two or more random variables, questions of **independence** are usually of great importance. In Example 25 we see that $f(x|y) = 2x$ does not depend on the given value $Y = y$, but this is clearly not the case in Example 24, where $f(x|y) = \frac{2x + 4y}{1 + 4y}$. Whenever the values of the conditional distribution of X given $Y = y$ do not depend on y , it follows that $f(x|y) = g(x)$, and hence the formulas of Definitions 12 and 13 yield

$$f(x, y) = f(x|y) \cdot h(y) = g(x) \cdot h(y)$$

That is, the values of the joint distribution are given by the products of the corresponding values of the two marginal distributions. Generalizing from this observation, let us now make the following definition.

DEFINITION 14. INDEPENDENCE OF DISCRETE RANDOM VARIABLES. *If $f(x_1, x_2, \dots, x_n)$ is the value of the joint probability distribution of the discrete random variables X_1, X_2, \dots, X_n at (x_1, x_2, \dots, x_n) and $f_i(x_i)$ is the value of the marginal distribution of X_i at x_i for $i = 1, 2, \dots, n$, then the n random variables are **independent** if and only if*

$$f(x_1, x_2, \dots, x_n) = f_1(x_1) \cdot f_2(x_2) \cdot \dots \cdot f_n(x_n)$$

for all (x_1, x_2, \dots, x_n) within their range.

To give a corresponding definition for continuous random variables, we simply substitute the word “density” for the word “distribution.”

With this definition of independence, it can easily be verified that the three random variables of Example 22 are not independent, but that the two random variables

X_1 and X_3 and also the two random variables X_2 and X_3 are **pairwise independent** (see Exercise 81).

The following examples serve to illustrate the use of Definition 14 in finding probabilities relating to several independent random variables.

EXAMPLE 26

Considering n independent flips of a balanced coin, let X_i be the number of heads (0 or 1) obtained in the i th flip for $i = 1, 2, \dots, n$. Find the joint probability distribution of these n random variables.

Solution

Since each of the random variables X_i , for $i = 1, 2, \dots, n$, has the probability distribution

$$f_i(x_i) = \frac{1}{2} \quad \text{for } x_i = 0, 1$$

and the n random variables are independent, their joint probability distribution is given by

$$\begin{aligned} f(x_1, x_2, \dots, x_n) &= f_1(x_1) \cdot f_2(x_2) \cdot \dots \cdot f_n(x_n) \\ &= \frac{1}{2} \cdot \frac{1}{2} \cdot \dots \cdot \frac{1}{2} = \left(\frac{1}{2}\right)^n \end{aligned}$$

where $x_i = 0$ or 1 for $i = 1, 2, \dots, n$.

EXAMPLE 27

Given the independent random variables X_1 , X_2 , and X_3 with the probability densities

$$\begin{aligned} f_1(x_1) &= \begin{cases} e^{-x_1} & \text{for } x_1 > 0 \\ 0 & \text{elsewhere} \end{cases} \\ f_2(x_2) &= \begin{cases} 2e^{-2x_2} & \text{for } x_2 > 0 \\ 0 & \text{elsewhere} \end{cases} \\ f_3(x_3) &= \begin{cases} 3e^{-3x_3} & \text{for } x_3 > 0 \\ 0 & \text{elsewhere} \end{cases} \end{aligned}$$

find their joint probability density, and use it to evaluate the probability $P(X_1 + X_2 \leq 1, X_3 > 1)$.

Solution

According to Definition 14, the values of the joint probability density are given by

$$\begin{aligned} f(x_1, x_2, x_3) &= f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_3) \\ &= e^{-x_1} \cdot 2e^{-2x_2} \cdot 3e^{-3x_3} \\ &= 6e^{-x_1 - 2x_2 - 3x_3} \end{aligned}$$

for $x_1 > 0, x_2 > 0, x_3 > 0$, and $f(x_1, x_2, x_3) = 0$ elsewhere. Thus,

$$\begin{aligned} P(X_1 + X_2 \leq 1, X_3 > 1) &= \int_1^\infty \int_0^1 \int_0^{1-x_2} 6e^{-x_1-2x_2-3x_3} dx_1 dx_2 dx_3 \\ &= (1 - 2e^{-1} + e^{-2})e^{-3} \\ &= 0.020 \end{aligned}$$

Exercises

69. Given the values of the joint probability distribution of X and Y shown in the table

		x	
		-1	1
y	-1	$\frac{1}{8}$	$\frac{1}{2}$
	0	0	$\frac{1}{4}$
	1	$\frac{1}{8}$	0

find

- (a) the marginal distribution of X ;
- (b) the marginal distribution of Y ;
- (c) the conditional distribution of X given $Y = -1$.

70. With reference to Exercise 42, find

- (a) the marginal distribution of X ;
- (b) the marginal distribution of Y ;
- (c) the conditional distribution of X given $Y = 1$;
- (d) the conditional distribution of Y given $X = 0$.

71. Given the joint probability distribution

$$f(x, y, z) = \frac{xyz}{108} \quad \text{for } x = 1, 2, 3; \quad y = 1, 2, 3; \quad z = 1, 2$$

find

- (a) the joint marginal distribution of X and Y ;
- (b) the joint marginal distribution of X and Z ;
- (c) the marginal distribution of X ;
- (d) the conditional distribution of Z given $X = 1$ and $Y = 2$;
- (e) the joint conditional distribution of Y and Z given $X = 3$.

72. With reference to Example 20, find

- (a) the **marginal distribution function** of X , that is, the function given by $G(x) = P(X \leq x)$ for $-\infty < x < \infty$;
- (b) the **conditional distribution function** of X given $Y = 1$, that is, the function given by $F(x|1) = P(X \leq x|Y = 1)$ for $-\infty < x < \infty$.

73. Check whether X and Y are independent if their joint probability distribution is given by

- (a) $f(x, y) = \frac{1}{4}$ for $x = -1$ and $y = -1, x = -1$ and $y = 1, x = 1$ and $y = -1$, and $x = 1$ and $y = 1$;
- (b) $f(x, y) = \frac{1}{3}$ for $x = 0$ and $y = 0, x = 0$ and $y = 1$, and $x = 1$ and $y = 1$.

74. If the joint probability density of X and Y is given by

$$f(x, y) = \begin{cases} \frac{1}{4}(2x + y) & \text{for } 0 < x < 1, 0 < y < 2 \\ 0 & \text{elsewhere} \end{cases}$$

find

- (a) the marginal density of X ;
- (b) the conditional density of Y given $X = \frac{1}{4}$.

75. With reference to Exercise 74, find

- (a) the marginal density of Y ;
- (b) the conditional density of X given $Y = 1$.

76. If the joint probability density of X and Y is given by

$$f(x, y) = \begin{cases} 24y(1 - x - y) & \text{for } x > 0, y > 0, x + y < 1 \\ 0 & \text{elsewhere} \end{cases}$$

find

- (a) the marginal density of X ;
- (b) the marginal density of Y .

Also determine whether the two random variables are independent.

77. With reference to Exercise 53, find

- (a) the marginal density of X ;
- (b) the marginal density of Y .

Also determine whether the two random variables are independent.

78. With reference to Example 22, find

- (a) the conditional density of X_2 given $X_1 = \frac{1}{3}$ and $X_3 = 2$;
- (b) the joint conditional density of X_2 and X_3 given $X_1 = \frac{1}{2}$.

79. If $F(x, y)$ is the value of the joint distribution function of X and Y at (x, y) , show that the **marginal distribution function** of X is given by

$$G(x) = F(x, \infty) \quad \text{for } -\infty < x < \infty$$

Use this result to find the marginal distribution function of X for the random variables of Exercise 54.

80. If $F(x_1, x_2, x_3)$ is the value of the joint distribution function of X_1, X_2 , and X_3 at (x_1, x_2, x_3) , show that the **joint marginal distribution function** of X_1 and X_3 is given by

$$M(x_1, x_3) = F(x_1, \infty, x_3) \quad \text{for } -\infty < x_1 < \infty, -\infty < x_3 < \infty$$

and that the **marginal distribution function** of X_1 is given by

$$G(x_1) = F(x_1, \infty, \infty) \quad \text{for } -\infty < x_1 < \infty$$

With reference to Example 19, use these results to find

- (a) the joint marginal distribution function of X_1 and X_3 ;
- (b) the marginal distribution function of X_1 .

81. With reference to Example 22, verify that the three random variables X_1, X_2 , and X_3 are not independent, but that the two random variables X_1 and X_3 and also the two random variables X_2 and X_3 are **pairwise independent**.

82. If the independent random variables X and Y have the marginal densities

$$f(x) = \begin{cases} \frac{1}{2} & \text{for } 0 < x < 2 \\ 0 & \text{elsewhere} \end{cases}$$

$$\pi(y) = \begin{cases} \frac{1}{3} & \text{for } 0 < y < 3 \\ 0 & \text{elsewhere} \end{cases}$$

find

- (a) the joint probability density of X and Y ;
- (b) the value of $P(X^2 + Y^2 > 1)$.

8 The Theory in Practice

This chapter has been about how probabilities can group themselves into probability distributions, and how, in the case of continuous random variables, these distributions become probability density functions. In practice, however, all data appear to be discrete. (Even if data arise from continuous random variables, the limitations of measuring instruments and roundoff produce discrete values.) In this section, we shall introduce some applications of the ideas of probability distributions and densities to the exploration of raw data, an important element of what is called **data analysis**.

When confronted with raw data, often consisting of a long list of measurements, it is difficult to understand what the data are informing us about the process, product, or service which gave rise to them. The following data, giving the response times of 30 integrated circuits (in picoseconds), illustrate this point:

<i>Integrated Circuit Response Times (ps)</i>									
4.6	4.0	3.7	4.1	4.1	5.6	4.5	6.0	6.0	3.4
3.4	4.6	3.7	4.2	4.6	4.7	4.1	3.7	3.4	3.3
3.7	4.1	4.5	4.6	4.4	4.8	4.3	4.4	5.1	3.9

Examination of this long list of numbers seems to tell us little other than, perhaps, the response times are greater than 3 ps or less than 7 ps. (If the list contained several hundred numbers, even this information would be difficult to elicit.)

A start at exploring data can be made by constructing a **stem-and-leaf display**. To construct such a display, the first digit of each response time is listed in a column at the left, and the associated second digits are listed to the right of each first digit. For the response-time data, we obtain the following stem-and-leaf display:

Probability Distributions and Probability Densities

3	7	4	4	7	7	4	3	7	9								
4	6	0	1	1	5	6	2	6	7	1	1	5	6	4	8	3	4
5	6	1															
6	0	0															

In this display, each row is a **stem** and the numbers in the column to the left of the vertical line are called **stem labels**. Each number on a stem to the right of the vertical line is called a **leaf**.

The stem-and-leaf display allows examination of the data in a way that would be difficult, if not impossible, from the original listing. For example, it can quickly be seen that there are more response times in the range 4.0 to 4.9 ps than any other, and that the great majority of circuits had response times of less than 5. This method of exploratory data analysis yields another advantage; namely there is no loss of information in a stem-and-leaf display.

The first two stems of this stem-and-leaf display contain the great majority of the observations, and more detail might be desirable. To obtain a finer subdivision of the data in each stem, a **double-stem display** can be constructed by dividing each stem in half so that the leaves in the first half of each stem are 0, 1, 2, 3, and 4, and those in the second half are 5, 6, 7, 8, and 9. The resulting double-stem display looks like this:

3f	4	4	4	3												
3s	7	7	7	7	9											
4f	0	1	1	2	1	1	4	3	4							
4s	6	5	6	6	7	5	6	8								
5*	6	1														
6*	0	0														

The stem labels include the letter *f* (for *first*) to denote that the leaves of this stem are 0–4, and *s* (for *second*) to denote that the leaves are 5–9. The asterisk is used with stem labels 5 and 6 to show that all 10 digits are included in these stems.

Numerical data can be grouped according to their values in several other ways in addition to stem-and-leaf displays.

DEFINITION 15. FREQUENCY DISTRIBUTION. *A grouping of numerical data into classes having definite upper and lower limits is called a **frequency distribution**.*

The construction of a frequency distribution is easily facilitated with a computer program such as MINITAB. The following discussion may be omitted if a computer program is used to construct frequency distributions.

To construct a frequency distribution, first a decision is made about the number of classes to use in grouping the data. The number of classes can be chosen to make the specification of upper and lower class limits convenient. Generally, the number of classes should increase as the number of observations becomes larger, but it is rarely helpful to use fewer than 5 or more than 15 classes.

The smallest and largest observations that can be put into each class are called the **class limits**. In choosing class limits, it is important that the classes do not overlap, so there is no ambiguity about which class contains any given observation. Also, enough classes should be included to accommodate all observations. Finally, the observations are tallied to determine the **class frequencies**, the number of observations falling into each class.

EXAMPLE 28

Construct a frequency distribution of the following compressive strengths (in psi) of concrete samples, given to the nearest 10 psi:

4890	4830	5490	4820	5230	4860	5040	5060	4500	5260
4610	5100	4730	5250	5540	4910	4430	4850	5040	5000
4600	4630	5330	5160	4950	4480	5310	4730	4700	4390
4710	5160	4970	4710	4430	4260	4890	5110	5030	4850
4820	4550	4970	4740	4840	4910	5200	4880	5150	4890
4900	4990	4570	4790	4480	5060	4340	4830	4670	4750

Solution

Since the smallest observation is 4260 and the largest is 5540, it will be convenient to choose seven classes, having the class limits 4200–4390, 4400–4590, . . . , 5400–5990. (Note that class limits of 4200–4400, 4400–4600, etc., are not used because they would overlap and assignment of 4400, for example, would be ambiguous; it could fit into either of the first two classes.) The following table exhibits the results of tallying the observations, that is, counting the number that fall into each class:

<i>Class Limits</i>	<i>Tally</i>	<i>Frequency</i>
4200–4390	///	3
4400–4590		7
4600–4790		12
4800–4990		19
5000–5190		11
5200–5390		6
5400–5590		2
	<i>Total</i>	60

Note the similarity between frequency distributions and probability distributions. A frequency distribution represents data, but a probability distribution represents a theoretical distribution of probabilities.

The midpoint between the upper class limit of a class and the lower class limit of the next class in a frequency distribution is called a **class boundary**. Class boundaries, rather than class marks, are used in constructing cumulative distributions (Exercise 88). The interval between successive class boundaries is called the **class interval**; it can also be defined as the difference between successive lower class limits or successive upper class limits. (Note that the class interval is *not* obtained by

subtracting the lower class limit of a class from its upper class limit.) A class can be represented by a single number, called the **class mark**. This number is calculated for any class by averaging its upper and lower class limits.

Once data have been grouped into a frequency distribution, each observation in a given class is treated as if its value is the class mark of that class. In so doing, its actual value is lost; it is known only that its value lies somewhere between the class limits of its class. Such an approximation is the price paid for the convenience of working with a frequency distribution.

EXAMPLE 29

For the frequency distribution of compressive strengths of concrete given in Example 28, find (a) the class boundaries, (b) the class interval, and (c) the class mark of each class.

Solution

- (a) The class boundaries of the first class are 4195–4395. The class boundaries of the second through the sixth classes are 4395–4595, 4595–4795, 4795–4995, 4995–5195, and 5195–5395, respectively. The class boundaries of the last class are 5395–5595. Note that the lower class boundary of the first class is calculated as if there were a class below the first class, and the upper class boundary of the last class is calculated as if there were a class above it. Also note that, unlike class limits, the class boundaries overlap.
- (b) The class interval is 200, the difference between the upper and lower class boundaries of any class. It also can be found by subtracting successive lower class limits, for example, $4400 - 4200 = 200$ psi, or by subtracting successive upper class limits, for example, $4590 - 4390 = 200$.
- (c) The class mark of the first class is $(4200 + 4390)/2 = 4295$; it is $(4400 + 4590)/2 = 4495$ for the second class; and the class marks are 4695, 4895, 5095, 5295, and 5495 for the remaining five classes. Note that the class interval, 200, also is given by the difference between any two successive class marks.

Histograms are easily constructed using most statistical software packages. Using MINITAB software to construct the histogram of compressive strengths, we obtain the result shown in Figure 13.

EXAMPLE 30

Suppose a wire is soldered to a board and pulled with continuously increasing force until the bond breaks. The forces required to break the solder bonds are as follows:

<i>Force Required to Break Solder Bonds (grams)</i>									
19.8	13.9	30.4	16.4	11.6	36.9	14.8	21.1	13.5	5.8
10.0	17.1	14.1	16.6	23.3	12.1	18.8	10.4	9.4	23.8
14.2	26.7	7.8	22.9	12.6	6.8	13.5	10.7	12.2	27.7
9.0	14.9	24.0	12.0	7.1	12.8	18.6	26.0	37.4	13.3

Use MINITAB or other statistical software to obtain a histogram of these data.

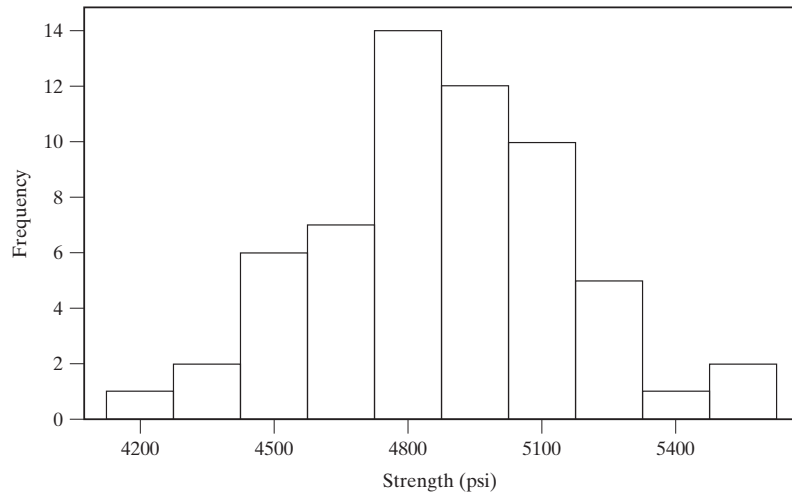


Figure 13. Histogram of compressive strengths.

Solution

The resulting histogram is shown in Figure 14. This histogram exhibits a right-hand “tail,” suggesting that while most of the solder bonds have low or moderate breaking strengths, a few had strengths that were much greater than the rest.

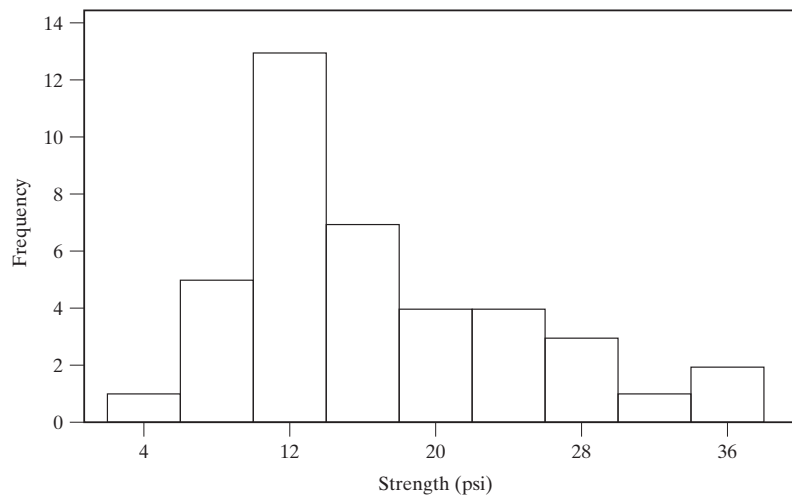


Figure 14. Histogram of solder-bond strengths.

Data having histograms with a long tail on the right or on the left are said to be **skewed**. A histogram exhibiting a long right-hand tail arises when the data have **positive skewness**. Likewise, if the tail is on the left, the data are said to have **negative skewness**. Examples of data that often are skewed include product lifetimes, many kinds of stress tests, workers’ incomes, and many weather-related phenomena, such as the proportion of cloudiness on a given day.

The shape of a histogram can be a valuable guide to a search for causes of production problems in the early stages of an investigation. For example, a skewed histogram often arises from “drifting” of machine settings from their nominal values. Sometimes skewed distributions do not arise from underlying causes but are the natural consequences of the type of measurements made. Some examples of “naturally” skewed data include the duration of telephone calls, the time intervals between emissions of radioactive particles, and, as previously mentioned, incomes of workers.

Histograms sometimes show more than one **mode**, or “high points.” A mode is a bar in a histogram that is surrounded by bars of lower frequency. A histogram exhibiting two modes is said to be **bimodal**, and one having more than two modes is said to be **multimodal**. An example of a bimodal histogram is shown in Figure 15. If there are several causes operating, each cause may generate its own distribution, and the histogram of all the data may be multimodal, each mode representing the center of the data that would arise from the corresponding cause if it were operating alone. Thus, multimodality can facilitate a search for underlying causes of error with the aim of eliminating them.

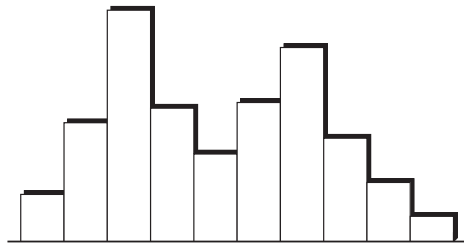


Figure 15. Bimodal histogram.

Applied Exercises

SECS. 1–2

83. With reference to Example 3, find the probability distribution of Y , the difference between the number of heads and the number of tails obtained in four tosses of a balanced coin.

84. An urn contains four balls numbered 1, 2, 3, and 4. If two balls are drawn from the urn at random (that is, each pair has the same chance of being selected) and Z is the sum of the numbers on the two balls drawn, find

(a) the probability distribution of Z and draw a histogram;

(b) the distribution function of Z and draw its graph.

85. A coin is biased so that heads is twice as likely as tails. For three independent tosses of the coin, find

(a) the probability distribution of X , the total number of heads;

(b) the probability of getting at most two heads.

86. With reference to Exercise 85, find the distribution function of the random variable X and plot its graph. Use the distribution function of X to find

(a) $P(1 < X \leq 3)$; (b) $P(X > 2)$.

87. The probability distribution of V , the weekly number of accidents at a certain intersection, is given by $g(0) = 0.40, g(1) = 0.30, g(2) = 0.20$, and $g(3) = 0.10$. Construct the distribution function of V and draw its graph.

88. With reference to Exercise 87, find the probability that there will be at least two accidents in any one week, using

(a) the original probabilities;

(b) the values of the distribution function.

89. This question has been intentionally omitted for this edition.

90. With reference to Exercise 80, find the distribution function of the sum of the spots on the dice, that is, the probability that this sum of the spots on the dice will be at most S , where $S = 2, 3, \dots, 12$.

SECS. 3–4

91. The actual amount of coffee (in grams) in a 230-gram jar filled by a certain machine is a random variable whose probability density is given by

$$f(x) = \begin{cases} 0 & \text{for } x \leq 227.5 \\ \frac{1}{5} & \text{for } 227.5 < x < 232.5 \\ 0 & \text{for } x \geq 232.5 \end{cases}$$

Find the probabilities that a 230-gram jar filled by this machine will contain

- (a) at most 228.65 grams of coffee;
- (b) anywhere from 229.34 to 231.66 grams of coffee;
- (c) at least 229.85 grams of coffee.

92. The number of minutes that a flight from Phoenix to Tucson is early or late is a random variable whose probability density is given by

$$f(x) = \begin{cases} \frac{1}{288}(36 - x^2) & \text{for } -6 < x < 6 \\ 0 & \text{elsewhere} \end{cases}$$

where negative values are indicative of the flight's being early and positive values are indicative of its being late. Find the probabilities that one of these flights will be

- (a) at least 2 minutes early;
- (b) at least 1 minute late;
- (c) anywhere from 1 to 3 minutes early;
- (d) exactly 5 minutes late.

93. The tread wear (in thousands of kilometers) that car owners get with a certain kind of tire is a random variable whose probability density is given by

$$f(x) = \begin{cases} \frac{1}{30}e^{-\frac{x}{30}} & \text{for } x > 0 \\ 0 & \text{for } x \leq 0 \end{cases}$$

Find the probabilities that one of these tires will last

- (a) at most 18,000 kilometers;
- (b) anywhere from 27,000 to 36,000 kilometers;
- (c) at least 48,000 kilometers.

94. The shelf life (in hours) of a certain perishable packaged food is a random variable whose probability density function is given by

$$f(x) = \begin{cases} \frac{20,000}{(x+100)^3} & \text{for } x > 0 \\ 0 & \text{elsewhere} \end{cases}$$

Find the probabilities that one of these packages will have a shelf life of

- (a) at least 200 hours;
- (b) at most 100 hours;
- (c) anywhere from 80 to 120 hours.

95. The total lifetime (in years) of five-year-old dogs of a certain breed is a random variable whose distribution function is given by

$$F(x) = \begin{cases} 0 & \text{for } x \leq 5 \\ 1 - \frac{25}{x^2} & \text{for } x > 5 \end{cases}$$

Find the probabilities that such a five-year-old dog will live

- (a) beyond 10 years;
- (b) less than eight years;
- (c) anywhere from 12 to 15 years.

96. In a certain city the daily consumption of water (in millions of liters) is a random variable whose probability density is given by

$$f(x) = \begin{cases} \frac{1}{9}xe^{-\frac{x}{3}} & \text{for } x > 0 \\ 0 & \text{elsewhere} \end{cases}$$

What are the probabilities that on a given day

- (a) the water consumption in this city is no more than 6 million liters;
- (b) the water supply is inadequate if the daily capacity of this city is 9 million liters?

SEC. 5

97. Two textbooks are selected at random from a shelf that contains three statistics texts, two mathematics texts, and three physics texts. If X is the number of statistics texts and Y the number of mathematics texts actually chosen, construct a table showing the values of the joint probability distribution of X and Y .

98. Suppose that we roll a pair of balanced dice and X is the number of dice that come up 1, and Y is the number of dice that come up 4, 5, or 6.

- (a) Draw a diagram like that of Figure 1 showing the values of X and Y associated with each of the 36 equally likely points of the sample space.
- (b) Construct a table showing the values of the joint probability distribution of X and Y .

99. If X is the number of heads and Y the number of heads minus the number of tails obtained in three flips of a balanced coin, construct a table showing the values of the joint probability distribution of X and Y .

100. A sharpshooter is aiming at a circular target with radius 1. If we draw a rectangular system of coordinates

with its origin at the center of the target, the coordinates of the point of impact, (X, Y) , are random variables having the joint probability density

$$f(x, y) = \begin{cases} \frac{1}{\pi} & \text{for } 0 < x^2 + y^2 < 1 \\ 0 & \text{elsewhere} \end{cases}$$

Find

- (a) $P[(X, Y) \in A]$, where A is the sector of the circle in the first quadrant bounded by the lines $y = 0$ and $y = x$;
- (b) $P[(X, Y) \in B]$, where $B = \{(x, y) | 0 < x^2 + y^2 < \frac{1}{2}\}$.

101. Suppose that P , the price of a certain commodity (in dollars), and S , its total sales (in 10,000 units), are random variables whose joint probability distribution can be approximated closely with the joint probability density

$$f(p, s) = \begin{cases} 5pe^{-ps} & \text{for } 0.20 < p < 0.40, s > 0 \\ 0 & \text{elsewhere} \end{cases}$$

Find the probabilities that

- (a) the price will be less than 30 cents and sales will exceed 20,000 units;
- (b) the price will be between 25 cents and 30 cents and sales will be less than 10,000 units.

102. A certain college gives aptitude tests in the sciences and the humanities to all entering freshmen. If X and Y are, respectively, the proportions of correct answers that a student gets on the tests in the two subjects, the joint probability distribution of these random variables can be approximated with the joint probability density

$$f(x, y) = \begin{cases} \frac{2}{5}(2x + 3y) & \text{for } 0 < x < 1, 0 < y < 1 \\ 0 & \text{elsewhere} \end{cases}$$

What are the probabilities that a student will get

- (a) less than 0.40 on both tests;
- (b) more than 0.80 on the science test and less than 0.50 on the humanities test?

SECS. 6-7

103. With reference to Exercise 97, find

- (a) the marginal distribution of X ;
- (b) the conditional distribution of Y given $X = 0$.

104. If X is the proportion of persons who will respond to one kind of mail-order solicitation, Y is the proportion of persons who will respond to another kind of mail-order

solicitation, and the joint probability density of X and Y is given by

$$f(x, y) = \begin{cases} \frac{2}{5}(x + 4y) & \text{for } 0 < x < 1, 0 < y < 1 \\ 0 & \text{elsewhere} \end{cases}$$

find the probabilities that

- (a) at least 30 percent will respond to the first kind of mail-order solicitation;
- (b) at most 50 percent will respond to the second kind of mail-order solicitation given that there has been a 20 percent response to the first kind of mail-order solicitation.

105. If two cards are randomly drawn (without replacement) from an ordinary deck of 52 playing cards, Z is the number of aces obtained in the first draw, and W is the total number of aces obtained in both draws, find

- (a) the joint probability distribution of Z and W ;
- (b) the marginal distribution of Z ;
- (c) the conditional distribution of W given $Z = 1$.

106. With reference to Exercise 101, find

- (a) the marginal density of P ;
- (b) the conditional density of S given $P = p$;
- (c) the probability that sales will be less than 30,000 units when $p = 25$ cents.

107. If X is the amount of money (in dollars) that a salesperson spends on gasoline during a day and Y is the corresponding amount of money (in dollars) for which he or she is reimbursed, the joint density of these two random variables is given by

$$f(x, y) = \begin{cases} \frac{1}{25} \left(\frac{20-x}{x} \right) & \text{for } 10 < x < 20, \frac{x}{2} < y < x \\ 0 & \text{elsewhere} \end{cases}$$

find

- (a) the marginal density of X ;
- (b) the conditional density of Y given $X = 12$;
- (c) the probability that the salesperson will be reimbursed at least \$8 when spending \$12.

108. Show that the two random variables of Exercise 102 are not independent.

109. The useful life (in hours) of a certain kind of integrated circuit is a random variable having the probability density

$$f(x) = \begin{cases} \frac{20,000}{(x + 100)^3} & \text{for } x > 0 \\ 0 & \text{elsewhere} \end{cases}$$

If three of these circuits operate independently, find
(a) the joint probability density of X_1, X_2 , and X_3 , representing the lengths of their useful lives;
(b) the value of $P(X_1 < 100, X_2 < 100, X_3 \geq 200)$.

SEC. 8

110. The following are the percentages of tin in measurements made on 24 solder joints:

61	63	59	54	65	60	62	61	67	60	55	68
57	64	65	62	59	59	60	62	61	63	58	61

(a) Construct a stem-and-leaf diagram using 5 and 6 as the stem labels.
(b) Construct a double-stem display.
(c) Which is more informative?

111. Suppose the first row of 12 observations in Exercise 110 came from solder connections made at station 105 and the second row came from station 107. Use a pair of stem-and-leaf diagrams to determine whether you should suspect a difference in the soldering process at the two stations.

112. Two different lathes turn shafts to be used in electric motors. Measurements made of their diameters (in cm) are

<i>Lathe A:</i> 1.42 1.38 1.40 1.41 1.39 1.44 1.36 1.42 1.40											
<i>Lathe B:</i> 1.47 1.31 1.56 1.33 1.29 1.46 1.28 1.51											

Construct two stem-and-leaf diagrams to see if you should suspect that the two lathes are turning out shafts of different diameters.

113. Use MINITAB or some other computer software to construct a stem-and-leaf display for the following data representing the time to make coke (in hours) in successive runs of a coking oven.

7.8	9.2	6.4	8.2	7.6	5.9	7.4	7.1	6.7	8.5
10.1	8.6	7.7	5.9	9.3	6.4	6.8	7.9	7.2	10.2
6.9	7.4	7.8	6.6	8.1	9.5	6.4	7.6	8.4	9.2

114. Use MINITAB or some other computer software to construct a stem-and-leaf display for the combined data of Exercise 112.

115. The following are the drying times (minutes) of 100 sheets coated with polyurethane under various ambient conditions:

45.6	50.3	55.1	63.0	58.2	65.5	51.1	57.4	60.4	54.9
56.1	62.1	43.5	63.8	64.9	59.9	63.0	67.7	53.8	57.9
61.8	52.2	61.2	51.6	58.6	73.8	53.9	64.1	57.2	75.4
55.9	70.1	46.2	63.6	56.0	48.1	62.2	58.8	50.8	68.1
51.4	73.9	66.7	42.9	71.0	56.1	60.8	58.6	70.6	62.2
59.9	47.5	72.5	62.0	56.8	54.3	61.0	66.3	52.6	63.5
64.3	63.6	53.5	55.1	62.8	63.3	64.7	54.9	54.4	69.6
64.2	59.3	60.6	57.1	68.3	46.7	73.7	56.8	62.9	58.4
68.5	68.9	62.1	62.8	74.4	43.8	40.0	64.4	50.8	49.9
55.8	66.8	67.0	64.8	57.6	68.3	42.5	64.4	48.3	56.5

Construct a frequency distribution of these data, using eight classes.

116. Eighty pilots were tested in a flight simulator and the time for each to take corrective action for a given emergency was measured in seconds, with the following results:

11.1	5.2	3.6	7.6	12.4	6.8	3.8	5.7	9.0	6.0	4.9	12.6
7.4	5.3	14.2	8.0	12.6	13.7	3.8	10.6	6.8	5.4	9.7	6.7
14.1	5.3	11.1	13.4	7.0	8.9	6.2	8.3	7.7	4.5	7.6	5.0
9.4	3.5	7.9	11.0	8.6	10.5	5.7	7.0	5.6	9.1	5.1	4.5
6.2	6.8	4.3	8.5	3.6	6.1	5.8	10.0	6.4	4.0	5.4	7.0
4.1	8.1	5.8	11.8	6.1	9.1	3.3	12.5	8.5	10.8	6.5	7.9
6.8	10.1	4.9	5.4	9.6	8.2	4.2	3.4				

Construct a frequency distribution of these data.

117. Find the class boundaries, the class interval, and the class marks of the frequency distribution constructed in Exercise 115.

118. Find the class boundaries, the class interval, and the class marks of the frequency distribution constructed in Exercise 116.

119. The following are the number of highway accidents reported on 30 successive days in a certain county:

6	4	0	3	5	6	2	0	0	12	3	7	2	1	1
0	4	0	0	0	1	8	0	2	4	7	3	6	2	0

Construct a frequency distribution of these data. Identify the class boundaries, the class marks, and the class interval.

120. A **percentage distribution** is obtained from a frequency distribution by replacing each frequency by 100 times the ratio of that frequency to the total frequency. Construct a percentage distribution using the reaction-time data of Exercise 116.

121. Construct a percentage distribution using the drying-time data of Exercise 115.

122. Percentage distributions are useful in comparing two frequency distributions having different total frequencies. Construct percentage distributions from the following two frequency distributions and determine whether the distributions of daily absences in the two departments follow similar patterns.

<i>FREQUENCIES</i>		
<i>Class Limits</i>	<i>Shipping Department</i>	<i>Security Department</i>
0–1	26	18
2–3	18	11
4–5	10	7
6–7	4	3
8–9	2	1
<i>Totals</i>	60	40

123. A **cumulative frequency distribution** is constructed from a frequency distribution by replacing each frequency with the sum of the frequency of the given class and the frequencies of all classes above it, and representing each class by its upper class boundary. Construct a cumulative frequency distribution using the data of Exercise 115.

124. Construct a cumulative frequency distribution using the data of Exercise 116.

125. Construct cumulative percentage distributions from the frequency distributions of absences given in Exercise 122.

126. Unequal class intervals. The small number of observations greater than 7 in Exercise 119 may cause some difficulty in constructing a frequency distribution. To keep class intervals equal, one is faced with the dilemma of either creating too many classes for only 30 observations or using a small number of classes with excessive loss of information in the first few classes. In such cases, one is tempted to drop the rule of equal-size classes, using a larger interval for the last class.

(a) If that were done, what would the resulting frequency distribution become?

(b) Is there a unique class interval?

127. The following are the times to failure of 38 light bulbs, given in hours of operation.

150	389	345	310	20	310	175	376	334	340
332	331	327	344	328	341	325	2	311	320
256	315	55	345	111	349	245	367	81	327
355	309	375	316	336	278	396	287		

(a) Dropping the rule that class intervals must be equal, construct a frequency distribution from these data.
(b) Can you find the class mark of every class?

128. (a) Construct a histogram of the reaction times of pilots from the data in Exercise 116.

(b) What can be said about the shape of this histogram?

129. (a) Construct a histogram of the drying times of polyurethane from the data in Exercise 115.

(b) What can be said about the shape of this histogram?

130. Use the data of Exercise 128 to illustrate that class marks are given by the midpoint between successive class boundaries as well as the midpoint between successive class limits.

131. Using the data of Exercise 129, show that the class marks also are given by the midpoint between successive class boundaries.

132. Construct a histogram using the solder-joint data in Exercise 110.

133. (a) Using only the first two rows of the data for the response times given in Section 8, construct a histogram.

(b) How would you describe the shape of the histogram?

134. (a) Combining the data for both lathes in Exercise 112, construct a histogram.

(b) How would you describe the shape of the histogram?

135. Use MINITAB or some other computer software to construct a histogram of the coking-time data given in Exercise 113.

136. Use MINITAB or some other computer software to construct a histogram of the drying-time data in Exercise 115.

137. A plot of the points (x, f) , where x represents the class mark of a given class in a frequency distribution and f represents its frequency, is called a **frequency polygon**. Construct a frequency polygon using the data in Exercise 116.

138. Construct a frequency polygon from the data in Exercise 115.

139. A plot of the cumulative frequency (see Exercise 123) on the y -axis and the corresponding upper class boundary on the x -axis is called an **ogive**.

- (a) Construct an ogive for the data of Exercise 115.
 (b) Using the same set of axes, relabel the y-axis so that the same graph also shows the ogive of the percentage distribution of drying times.

140. (a) Construct an ogive for the reaction times given in Exercise 116.
 (b) Construct an ogive representing the cumulative percentage distribution.

References

More advanced or more detailed treatments of the material in this chapter may be found in
 BRUNK, H. D., *An Introduction to Mathematical Statistics*, 3rd ed. Lexington, Mass.: Xerox College Publishing, 1975,
 DEGROOT, M. H., *Probability and Statistics*, 2nd ed. Reading, Mass.: Addison-Wesley Publishing Company, Inc., 1986,
 FRASER, D. A. S., *Probability and Statistics: Theory and Applications*. North Scituate, Mass.: Duxbury Press, 1976,

HOGG, R. V., and CRAIG, A. T., *Introduction to Mathematical Statistics*, 4th ed. New York: Macmillan Publishing Co., Inc., 1978,
 KENDALL, M. G., and STUART, A., *The Advanced Theory of Statistics*, Vol. 1, 4th ed. New York: Macmillan Publishing Co., Inc., 1977,
 KHAZANIE, R., *Basic Probability Theory and Applications*. Pacific Palisades, Calif.: Goodyear Publishing Company, Inc., 1976.

Answers to Odd-Numbered Exercises

- 1 (a) no, because $f(4)$ is negative; (b) yes; (c) no, because the sum of the probabilities is less than 1.
 5 $0 < k < 1$.
 9 (a) no, because $F(4)$ exceeds 1; (b) no, because $F(2)$ is less than $F(1)$; (c) yes.
 11 (a) $\frac{1}{2}$; (b) $\frac{1}{6}$; (c) $f(1) = \frac{1}{3}$; $f(4) = \frac{1}{6}$; $f(6) = \frac{2}{3}$; $f(10) = \frac{1}{6}$; $f(x) = 0$ elsewhere.
 13 (a) $\frac{3}{4}$; (b) $\frac{1}{4}$; (c) $\frac{1}{2}$; (d) $\frac{3}{4}$; (e) $\frac{1}{2}$; (f) $\frac{1}{4}$.
 17 (b) $\frac{2}{5}$.
 19 (c) 0.124.

$$21 F(y) = \begin{cases} 0 & \text{for } y \leq 2 \\ \frac{1}{16}(y^2 + 2y - 8) & \text{for } 2 < y < 4 \\ 1 & \text{for } y \geq 4 \end{cases}$$

The probabilities are 0.454 and 0.1519.

$$23 F(x) = \begin{cases} 0 & \text{for } x \leq 0 \\ \frac{1}{2}\sqrt{x} & \text{for } 0 < x < 4; \text{ (b) } \frac{1}{4} \text{ and } \frac{1}{2}. \\ 1 & \text{for } x \geq 4 \end{cases}$$

$$25 F(z) = \begin{cases} 0 & \text{for } z \leq 0 \\ 1 - e^{-z^2} & \text{for } z > 0 \end{cases}$$

$$27 G(x) = \begin{cases} 0 & \text{for } x \leq 0 \\ 3x^2 - 2x^3 & \text{for } 0 < x < 1 \\ 1 & \text{for } x \geq 1 \end{cases}$$

The probabilities are $\frac{5}{32}$ and $\frac{1}{2}$.

$$29 F(x) = \begin{cases} 0 & \text{for } x \leq 0 \\ \frac{x^2}{2} & \text{for } 0 < x < 1 \\ 2x - \frac{x^2}{2} - 1 & \text{for } 1 \leq x < 2 \\ 1 & \text{for } x \geq 2 \end{cases}$$

$$31 F(x) = \begin{cases} 0 & \text{for } x \leq 0 \\ \frac{x^2}{4} & \text{for } 0 < x \leq 1 \\ \frac{1}{4}(2x - 1) & \text{for } 1 < x \leq 2 \\ \frac{1}{4}(6x - x^2 - 5) & \text{for } 2 < x < 3 \\ 1 & \text{for } x \geq 3 \end{cases}$$

$$33 f(x) = \frac{1}{2} \text{ for } -1 < x < 1 \text{ and } f(x) = 0 \text{ elsewhere.}$$

$$35 f(y) = \frac{18}{y^3} \text{ for } y > 0 \text{ and } f(y) = 0 \text{ elsewhere; the two probabilities are } \frac{16}{25} \text{ and } \frac{9}{64}.$$

$$37 \text{ The three probabilities are } 1 - 3e^{-2}, 2e^{-1} - 4e^{-3}, \text{ and } 5e^{-5}.$$

$$39 \text{ (a) } F(x) = 0; \text{ (b) } F(x) = \frac{1}{2}x; \text{ (c) } F(x) = \frac{1}{2}(x+1); \text{ (d) } F(x) = 0.$$

$$41 \text{ The probabilities are } \frac{1}{4}, \frac{1}{4}, \frac{3}{8}, \text{ and } \frac{1}{2}.$$

$$43 \text{ (a) } \frac{1}{4}; \text{ (b) } 0; \text{ (c) } \frac{7}{24}; \text{ (d) } \frac{119}{120}.$$

$$45 \text{ (a) } \frac{29}{89}; \text{ (b) } \frac{5}{89}; \text{ (c) } \frac{55}{89}.$$

Probability Distributions and Probability Densities

47

		x			
		0	1	2	3
y	0	0	$\frac{1}{30}$	$\frac{1}{10}$	$\frac{1}{5}$
	1	$\frac{1}{30}$	$\frac{2}{15}$	$\frac{3}{10}$	$\frac{8}{15}$
	2	$\frac{1}{10}$	$\frac{3}{10}$	$\frac{3}{5}$	1

49 $k = 2$.

51 (a) $\frac{1}{2}$; (b) $\frac{5}{9}$; (c) $\frac{1}{3}$.

53 $1 - \frac{1}{2} \ln 2 = 0.6354$.

55 $(e^{-1} - e^{-4})^2$.

57 $(e^{-2} - e^{-3})^2$.

63 (a) $\frac{1}{18}$; (b) $\frac{7}{27}$.

65 $k = 144$.

71 (a) $m(x, y) = \frac{xy}{36}$ for $x = 1, 2, 3$ and $y = 1, 2, 3$;

(b) $n(x, z) = \frac{xz}{18}$ for $x = 1, 2, 3$ and $z = 1, 2$;

(c) $g(x) = \frac{x}{6}$ for $x = 1, 2, 3$; (d) $\phi(z|1, 2) = \frac{z}{3}$ for $z = 1, 2$;

(e) $\psi(y, z|3) = \frac{yz}{18}$ for $y = 1, 2, 3$ and $z = 1, 2$.

73 (a) Independent; (b) not independent.

75 (a) $h(y) = \frac{1}{4}(1 + y)$ for $0 < y < 2$ and $h(y) = 0$ elsewhere;

(b) $f(x|1) = \frac{1}{2}(2x + 1)$ for $0 < x < 1$ and $f(x|1) = 0$ elsewhere.

77 (a) $g(x) = -\ln x$ for $0 < x < 1$ and $g(x) = 0$ elsewhere;

(b) $h(y) = 1$ for $0 < y < 1$ and $h(y) = 0$ elsewhere. The two random variables are not independent.

79 $G(x) = 1 - e^{-x^2}$ for $x > 0$ and $G(x) = 0$ elsewhere.

83

Y	-4	-2	0	2	4
P(Y)	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$

85 (a)

X	0	1	2	3
P(X)	$\frac{1}{27}$	$\frac{6}{27}$	$\frac{12}{27}$	$\frac{8}{27}$

(b) $\frac{19}{27}$.

$$87 F(V) = \begin{cases} 0 & \text{for } V < 0 \\ 0.40 & \text{for } 0 \leq V < 1 \\ 0.70 & \text{for } 1 \leq V < 2 \\ 0.90 & \text{for } 2 \leq V < 3 \\ 1 & \text{for } V \geq 3 \end{cases}$$

89 Yes; $\sum_{x=2}^{12} f(x) = 1$.

91 (a) 0.23; (b) 0.464; (c) 0.53.

93 (a) 0.4512; (b) 0.1054; (c) 0.2019.

95 (a) $\frac{1}{4}$; (b) $\frac{39}{64}$; (c) $\frac{1}{16}$.

101 (a) 0.3038; (b) $\frac{1}{221}$.

103 (a) $g(0) = \frac{5}{14}$, $g(1) = \frac{15}{28}$, $g(2) = \frac{3}{28}$;

(b) $\phi(0|0) = \frac{3}{10}$, $\phi(1|0) = \frac{6}{10}$, $\phi(2|0) = \frac{1}{10}$.

105 (a) $f(0, 0) = \frac{188}{221}$, $f(0, 1) = \frac{16}{221}$, $f(1, 0) = \frac{16}{221}$, $f(1, 1) = \frac{1}{221}$.

(b) $g(0) = \frac{204}{221}$, $g(1) = \frac{17}{221}$; (c) $\phi(0|0) = \frac{16}{17}$, $\phi(1, 1) = \frac{1}{17}$.

107 (a) $g(x) = \frac{20-x}{50}$ for $10 < x < 20$ and $g(x) = 0$ elsewhere; (b) $\phi(y|12) = \frac{1}{6}$ for $6 < y < 12$ and $\phi(y|12) = 0$ elsewhere; (c) $\frac{2}{3}$.

109 (a) $f(x_1, x_2, x_3) = \frac{(20,000)^3}{(x_1 + 100)^3(x_2 + 100)^3(x_3 + 100)^3}$ for $x_1 > 0$, $x_2 > 0$, $x_3 > 0$ and $f(x_1, x_2, x_3) = 0$ elsewhere; (b) $\frac{1}{16}$.

111 Station 107 data show less variability than station 105 data.