[09-04-22-MAM11] Probability (first of two chapters

Modern Advanced Mathematics

By

ARTHUR W. WEEKS

JACKSON B. ADKINS

and

RANSOM V. LYNCH

The Phillips Exeter Academy
EXETER, NEW HAMPSHIRE

GINN AND COMPANY
A XEROX COMPANY

15

Probability

15.1 Sample Spaces and Events

There is an old saying that nothing in this life is certain but death and taxes. Though the saying may not be literally true, it does contain an important element of truth: we must often make decisions and take action with no certain knowledge of what the outcome will be. Life is full of uncertainty, and the mathematics of the uncertain is called probability.

Historically, this branch of mathematics grew out of the study of certain games of chance, and the student will find that many of the problems of elementary probability still are concerned with such activities as tossing coins, throwing dice, and dealing cards. The important modern applications of probability, however, are to more respectable activities: computing life insurance premiums, controlling the quality of mass-produced items, planning telephone networks, forecasting the weather, growing better chickens, evaluating medical treatments, deciding who wrote historical documents, and many others.

A throw of a coin is unpredictable: we have no way of knowing whether it will fall heads or fall tails. If we toss the coin repeatedly, there is no apparent order in the pattern of heads and tails, and each throw is as unpredictable as its predecessors. Yet, in all this uncertainty and irregularity, a certain regularity can be found: if we toss the coin many times, about half the tosses will show heads. We often express our belief in this regularity by saying, "There is a fifty-fifty chance of heads," or "The probability of heads is $\frac{1}{2}$."

A throw of heads with a coin is an example of a random event: we have no way of knowing whether or not it will occur on a single trial, but the long-run proportion of heads approaches a fixed limiting value. It is with random events that we shall concern ourselves in this study of probability. We shall therefore be concerned with experiments such as tossing a coin, rolling a pair of dice, or measuring the length of a metal rod. Each of these experiments can be repeated many times under essentially identical circumstances, so that it makes sense to talk about the long-run frequencies, or proportions, of the various possible outcomes. Since we shall confine our attention to experiments of this sort, we shall be unable to make any estimate of the probability that Bacon wrote the works attributed to Shakespeare, or that there is life on Mars, or that the moon is made of green cheese.

The first step in analyzing an experiment is to describe the set of possible outcomes of the experiment. This set is called a *sample space* for the experiment, and each possible outcome is often called a *point* in the sample space, or a *sample point*. In most cases, an experiment can be analyzed in more than one way, so that there may be more than one possible sample space for the same experiment. In a roll of a die, for example, one of the six numbers on the die must fall uppermost, and we may therefore take as a sample space for this experiment $\{1, 2, 3, 4, 5, 6\}$. Another possible sample space for the same experiment, however, is {odd number, even number}, and still another is {1, 3, 5, even number}. Each of these sets has the essential features that (i) each element of the set corresponds to a possible outcome of the experiment, and (ii) each possible outcome of the experiment corresponds to one and only one element of the set. On the other hand, $\{1, 2, 3, 4, 5, 6, 7\}$ is not an acceptable sample space for this experiment because "7" is not a possible outcome, {1, odd number, even number} is not acceptable because a throw of 1 corresponds to two elements of the set, and {3, 5, even number} is not acceptable because a throw of 1 corresponds to no element of the set.

Similarly, if the experiment consists in tossing a penny and a nickel, then a possible sample space is $S_1 = \{HH, HT, TH, TT\}$, where the first letter in each pair indicates whether the penny showed heads (H) or tails (T), and the second letter gives the same information about the nickel. Another sample space for this experiment is $S_2 =$ {0 heads, 1 head, 2 heads}. Each of these is acceptable, but S_1 represents a more detailed analysis of the experiment. If we know an outcome in S_1, we can determine the corresponding outcome in S_2, but the converse is false, since the outcome "1 head" may correspond to either HT or TH.

We summarize the preceding discussion in a definition.

▶ **Definition.** **A set *S* is a sample space for an experiment if and only if (*i*) each element of *S* corresponds to a possible outcome of the experiment, and (*ii*) each possible outcome of the experiment corresponds to one and only one element of *S*.**

We also require a second definition.

▶ **Definition.** **An event is a subset of a sample space.**

Many events have simple descriptions in ordinary, non-technical language. In the sample space S_1 for the two-coin experiment, for example, the event "One head" is the subset $\{HT, TH\}$, the event "At least one head" is $\{HH, HT, TH\}$, the event "Both coins the same" is $\{HH, TT\}$, and so on. If, when an experiment is performed, the outcome is an element of the event E, we then say that the event E has occurred.

Example. Three chips, numbered 1, 2, 3, are placed in a bowl. A chip is drawn and replaced, and a second drawing is made. A sample space S for this experiment is the set of ordered pairs (x, y) in which x is the number obtained on the first draw and y is the number obtained on the second draw. It is convenient to display the elements of this set in a table:

$x \backslash y$	1	2	3
1	(1, 1)	(1, 2)	(1, 3)
2	(2, 1)	(2, 2)	(2, 3)
3	(3, 1)	(3, 2)	(3, 3)

We may list several events in this sample space: E_1 = "Both numbers are the same" $= \{(x, y): x = y\} = \{(1, 1), (2, 2), (3, 3)\}$; E_2 = "The first number is greater than the second" $= \{(x, y): x > y\} = \{(2, 1), (3, 1), (3, 2)\}$; E_3 = "The sum of the numbers is 4" $= \{(x, y): x + y = 4\} = \{(3, 1), (2, 2), (1, 3)\}$; and so on. If we actually perform the experiment and obtain the outcome (3, 1), then, since $(3, 1) \in E_2$, we say that E_2 has occurred, and, since $(3, 1) \in E_3$, we also say that E_3 has occurred. It follows, of course, that the event $E_2 \cap E_3$ has occurred.

Exercises [A]

1. An experiment consists in selecting a letter from the word *F A C E-T I O U S*. Which of the following are acceptable sample spaces for this experiment?
   (a) {consonant, vowel}
   (b) {vowel, F, C, T, S}
   (c) {vowel, F, C, T, S, X}
   (d) {consonant, A, E, O, U}
   (e) {consonant, A, C, E, I, O, U}
   (f) {C, a letter coming before C, a letter coming after C}

2. Describe two sample spaces for each of the following experiments. In each case, determine the number of points in the sample space.
   (a) A coin is tossed 3 times. (b) A coin is tossed n times.
   (c) A card is dealt from an ordinary deck.
   (d) A two-card hand is dealt from an ordinary deck.
   (e) Three letters are selected, with replacement, from the English alphabet.

3. The experiment described in the example on page 501 is modified in that the first chip is not replaced before the second drawing.
   (a) Describe a sample space of ordered pairs (x, y) for this experiment.
   (b) List the elements of the following events.
      (i) $E_1 = \{(x, y) : x > y\}$ (ii) $E_2 = \{(x, y) : x = 3\}$
      (iii) $E_3 = \{(x, y) : y = 2\}$ (iv) $E_4 = \{(x, y) : x + y = 4\}$
      (v) $E_1 \cup E_2$ (vi) $E_1 \cap E_2$
      (vii) $E_3 \cup E_4$ (viii) $E_3 \cap E_4$
      (ix) $E_2 \cap E_3$ (x) $E_1 \cap E_3$

4. A red die and a green die are thrown. Make a table showing the elements of the sample space of ordered pairs $(x, y) =$ (number on red die, number on green die). (It would be a good idea to make this table on a separate card that you can keep in your book, as we shall make frequent reference to this sample space in what follows.) List the elements of the following events.
   (a) $E_1 = \{(x, y) : x + y = 5\}$ (b) $E_2 = \{(x, y) : x + y = 7\}$
   (c) $E_3 = \{(x, y) : x = 4\}$ (d) $E_4 =$ "Both numbers the same"
   (e) $E_5 = \{(x, y) : x > 4, y < 3\}$ (f) $E_2 \cap E_4$
   (g) $E_1 \cap E_4$ (h) $\overline{E_3} \cap E_2$
   (i) $E_3 \cap \overline{E_4}$ (j) $E_2 \cup E_3$

15.2 Equally Likely Outcomes

If a coin is tossed repeatedly, there are intuitive, common-sense reasons to believe that heads will come up about half the time. This belief is often expressed in the assertion that, on any one toss, the probability of heads is $\frac{1}{2}$. If an ordinary die is rolled repeatedly, the symmetry of the die suggests that, in the long run, a particular face—say the face numbered 2—will turn up about $\frac{1}{6}$ of the time. We express our judgment of this situation by saying that, on any one throw, the probability of a 2 is $\frac{1}{6}$.

This line of reasoning is easily generalized. If an experiment has N possible outcomes, and if there is a natural symmetry among these outcomes suggesting that they are all equally likely, we then say that the probability of any one of them is $\frac{1}{N}\cdot$ There are 2 equally likely outcomes in tossing a coin; hence the probability of either is $\frac{1}{2}$. There are 6 equally likely outcomes in rolling a die; hence the probability of any one of them is $\frac{1}{6}$.

The argument can be extended. In the sample space $S = \{1, 2, 3, 4, 5, 6\}$ for the experiment of rolling a die, the event "Outcome greater than 2" is $E = \{3, 4, 5, 6\}$. If the die is rolled repeatedly, we expect that, in the long run, each outcome in S will occur about $\frac{1}{6}$ of the time. But 4 of these outcomes are in E, so that we can expect E to oocur about $\frac{4}{6}$, or $\frac{2}{3}$, of the time. We therefore conclude that the probability of the event E on a single roll of the die is $\frac{2}{3}$. We note that this probability is $\frac{n(E)}{n(S)}$, the ratio of the number of outcomes in E to the total number of outcomes in the whole sample space S.

If the outcome o of an experiment is an element of an event E, then the classical language of probability (much of it dating from the seventeenth century) calls o an outcome favorable to E. The classical statement of our preceding conclusion is then this: the probability of an event is the number of outcomes favorable to the event divided by the total number of possible outcomes. In the language of Chapter 14, this classical definition may be stated thus:

▶ **Definition.** **If all outcomes in a sample space S are equally likely, then the probability of any event $A \subseteq S$ is**

$$P(A) = \frac{n(A)}{n(S)}, \tag{1}$$

where $n(A)$, $n(S)$ are respectively the number of outcomes in A and the number in S.

This definition has two major flaws. First, the decision as to whether all possible outcomes are equally likely is a subjective one, on which there may be disagreement. Second, there are many experiments, such as tossing a bent coin or a loaded (weighted or asymmetrical) die, in which it would be unrealistic to consider all possible outcomes as equally likely. We shall see in the next section how these defects can be remedied. The classical definition (*1*) suffices, however, for the solution of an extraordinary variety of problems. We shall examine some of these problems in this section.

Example 1. Two numbers are selected, without replacement, from $\{1, 2, 3, 4, 5\}$. What is the probability that (a) the numbers selected are 1 and 4? (b) one of the numbers selected is 3?

Solution: The nature of the experiment suggests that we take for our sample space the set of $\binom{5}{2} = 10$ two-element subsets of the given set. For reference, we list these subsets.

$$\begin{matrix} \{1,2\} & \{1,3\} & \{1,4\} & \{1,5\} \\ & \{2,3\} & \{2,4\} & \{2,5\} \\ & & \{3,4\} & \{3,5\} \\ & & & \{4,5\} \end{matrix}$$

If the selection is done without bias, we can assume that all 10 possible outcomes are equally likely. (a) Since $\{1, 4\}$ is a single point in S, we have $n(\{1, 4\}) = 1$. Thus, $P(\{1, 4\}) = \frac{n(\{1, 4\})}{n(S)} = \frac{1}{10}$, by (*1*). (b) Four outcomes, namely $\{1, 3\}$, $\{2, 3\}$, $\{3, 4\}$, $\{3, 5\}$, are elements of the event "One of the numbers is 3." Therefore, $P(\text{One of the numbers is 3}) = \frac{4}{10} = \frac{2}{5}$.

Example 2. What is the probability that a 5-card hand contains no pair?

Solution: We know from the example on page 491, that there are $\binom{52}{5} = 2{,}598{,}960$ 5-card hands. These are the elements of the appropriate sample space, the set of all possible outcomes of the experiment "deal a 5-card hand." We know from Example 1 on page 492, that there are $\binom{13}{5} \cdot 4^5 = 1{,}317{,}888$ of these hands that do not contain a pair. Hence, $P(\text{No pair}) = \frac{1{,}317{,}888}{2{,}598{,}960} \approx 0.507$.

As the notation $P(A)$ suggests, P is the name of a function, the probability function. In any one instance, the domain of P is the set of all events in some sample space S, and its range, as we shall prove, is some subset of the interval $[0, 1]$. Because equation (*1*) reduces the computation of $P(A)$ to a matter of counting and then dividing, it follows that the function P so defined has many properties much like properties of the counting function n. Most of these properties are contained in the following theorems and corollaries. In each of them, A, B, and so on, are events in a finite sample space S. We assume that S itself is not empty, that is, that $n(S) > 0$, and that all outcomes in S are equally likely.

▶ **Theorem 1.** $A \cap B = \emptyset \longrightarrow P(A \cup B) = P(A) + P(B)$.

Proof: This theorem is analogous to property (*1*) of the counting function n, stated on page 473. By that property, we have

$$n(A \cup B) = n(A) + n(B).$$

Dividing each term of the equation by $n(S)$, we have

$$\frac{n(A \cup B)}{n(S)} = \frac{n(A)}{n(S)} + \frac{n(B)}{n(S)},$$

and, by the definition of probability, this reduces to

$$P(A \cup B) = P(A) + P(B).$$

As has been noted (page 501), if an experiment is performed and the outcome o is an element of an event E, we then say that E has occurred. If $A \subseteq S$, $B \subseteq S$, and $A \cap B = \emptyset$, then no outcome in S belongs to both A and B, and it is therefore impossible, in any one trial of the experiment, that both A and B should occur. If A occurs, then B does not: $o \in A \longrightarrow o \notin B$. Similarly, $o \in B \longrightarrow o \notin A$. For this reason, we call such events *mutually exclusive.*

▶ **Definition.** Events A, B are mutually exclusive if and only if $A \cap B = \emptyset$.

Using this terminology, the preceding theorem can be restated as follows.

▶ **Theorem 1.** If events A and B are mutually exclusive, then the probability that A or B occurs is the sum of the probability that A occurs and the probability that B occurs.

Example. A die is rolled. What is the probability of a 3 or a 5?

Solution: The appropriate sample space, the set of equally likely outcomes, is $\{1, 2, 3, 4, 5, 6\}$. Let $E = \{3\}, F = \{5\}$. Then $E \cap F = \emptyset$, and therefore $P(E \cup F) = P(E) + P(F) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$, by Theorem 1.

If an event is empty, then the event does not occur, whatever may be the outcome of the corresponding experiment. It is therefore reasonable that the probability of the empty event be zero. There are various ways of proving that this is so. One of them is a corollary to Theorem 1.

▶ **Corollary 1.** $P(\emptyset) = 0$.

Proof: (See Exercise 2(a), page 484.) If $A = B = \emptyset$, then $A \cap B = A \cup B = \emptyset$, and Theorem 1 becomes

$$P(\emptyset) = P(\emptyset) + P(\emptyset).$$

Subtracting $P(\emptyset)$ from both sides gives $P(\emptyset) = 0$.

If three or more events are such that any two of them are mutually exclusive, then they are said to be *mutually exclusive in pairs.* Theorem 1 can be extended to the union of any collection of such events.

▶ **Theorem 2.** Let $A_1, A_2, \ldots, A_k$ be events such that $i \neq j \longrightarrow A_i \cap A_j = \emptyset$. Then

$$P(A_1 \cup A_2 \cup \cdots \cup A_k) = \sum_{i=1}^{k} P(A_i).$$

Proof: The proof starts with Theorem 1 of Chapter 14 (page 474), and proceeds essentially as in the proof of Theorem 1 of this chapter. The details are left to the student as an exercise.

Example. In the experiment of the preceding example, what is the probability of rolling 3 or 4 or 5 or 6?

Solution: Let $E_1 = \{3\}$, $E_2 = \{4\}$, $E_3 = \{5\}$, $E_4 = \{6\}$. Then these events are mutually exclusive in pairs, and therefore $P(E_1 \cup E_2 \cup E_3 \cup E_4) = P(E_1) + P(E_2) + P(E_3) + P(E_4) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{2}{3}$, by Theorem 2.

The probability of any event can be inferred from the probability of the complementary event. This fact is often helpful in the computation of probabilities. Specifically, we have the following.

▶ **Theorem 3.** $P(\bar{A}) = 1 - P(A)$.

Proof: Use the result in Theorem 2 of Chapter 14 (page 475), and divide by $n(S) = N$. The details are left to the student as an exercise.

Example. What is the probability that a 5-card hand contains a pair?

Solution: The event "Pair" is the complement of the event "No pair." We found in Example 2 on page 504 that $P(\text{No pair}) = \frac{1{,}317{,}888}{2{,}598{,}960}$. Hence,

$$P(\text{Pair}) = 1 - \frac{1{,}317{,}888}{2{,}598{,}960} = \frac{1{,}281{,}072}{2{,}598{,}960} \approx 0.493.$$

Each possible outcome of an experiment is an element of any sample space S for that experiment by definition of sample space. No matter what the outcome, then, the event S occurs; it is certain to happen. The corresponding probability is 1. This property may be proved as a corollary to Theorem 3.

▶ **Corollary 2.** $P(S) = 1$.

Proof: Since $\bar{\emptyset} = S$ and $P(\emptyset) = 0$ by Corollary 1, if we put $\emptyset$ for A in Theorem 3, we have

$$P(\bar{\emptyset}) = 1 - P(\emptyset),$$

or

$$P(S) = 1 - 0 = 1.$$

If an event A is a subevent of an event B, that is, if $A \subseteq B$, then every outcome in A is also an outcome in B. Hence, if A occurs, then B also occurs, and we may say that the occurrence of A implies the occurrence of B. Thus, if $A \subseteq B$, then the probability of B is not less than the probability of A.

▶ **Theorem 4.** $A \subseteq B \longrightarrow P(A) \leq P(B)$.

Proof: This theorem follows directly from Theorem 3 of Chapter 14 (page 478). The details are left to the student.

We have now established the essential features of the scale on which probabilities are measured. The results are summarized in the following theorem which we state as a corollary to Theorem 4.

▶ **Corollary 3.** For any event A, $0 \leq P(A) \leq 1$.

Proof: For any event A, $\emptyset \subseteq A \subseteq S$. The rest of the proof is left to the student as an exercise.

It follows from Theorem 1 and Corollary 3 that the probability function P is a measure, as defined on page 476. We complete our list of the properties of this function with a theorem that tells us how to find the probability of the union of any two events, whether or not they are mutually exclusive.

▶ **Theorem 5.** $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Proof: Since P is a measure, this is just a special case of Exercise 2(c), page 484. Alternatively, a proof may be based on Theorem 4 of Chapter 14 (page 479). The details are left to the student as an exercise.

Example. In the sample space of Example 1 on page 504, let $G = \{1, 3, 5\}$, $H = \{3, 4, 5, 6\}$, and compute $P(G \cup H)$.

Solution: $G \cap H = \{3, 5\}$, so that $n(G \cap H) = 2$, and therefore $P(G \cap H) = \frac{2}{6}$. $P(G) = \frac{3}{6}$, $P(H) = \frac{4}{6}$. Hence, $P(G \cup H) = P(G) + P(H) - P(G \cap H) = \frac{3}{6} + \frac{4}{6} - \frac{2}{6} = \frac{5}{6}$, by Theorem 5. We can check this result by observing that $G \cup H = \{1, 3, 4, 5, 6\}$, so that $n(G \cup H) = 5$, and therefore $P(G \cup H) = \frac{5}{6}$ by (*1*).

One often hears or reads about *odds* for or against an event, such as the victory of one or the other team in an athletic contest. There is a simple relationship between odds and probabilities. If the probability of an event A is $P(A) = p$ and the probability of the complementary event $\overline{A}$ is $P(\overline{A}) = q$, then odds for A are $p : q$ (read "p to q"), and odds against A (and for $\overline{A}$) are $q : p$. The ratio $p : q$, which is equivalent to the fraction $\frac{p}{q}$, is usually simplified if possible. For example, if $p = 0.7$, then we know from Theorem 3 that $q = 1 - p = 0.3$. Hence odds for A are $0.7 : 0.3$, which may be simplified to $7 : 3$, and odds against A are $0.3 : 0.7$, or $3 : 7$.

If odds for A are $a : b$, and if $P(A) = p$, $P(\overline{A}) = q = 1 - p$, then odds for A are also $p : q = p : (1 - p)$. Expressing these odds as fractions, we have

$$\frac{a}{b} = \frac{p}{1-p},$$

and solving for p in terms of a and b, we find

$$p = \frac{a}{a+b}.$$

Thus, if odds for A are 3 : 5, then $P(A) = \dfrac{3}{3+5} = \dfrac{3}{8}$. We summarize this discussion in the following definition.

▶ **Definition.** Odds for an event A are $a : b$, and odds against A (and for $\overline{A}$) are $b : a$, if and only if

$$P(A) = \frac{a}{a+b}.$$

Exercises [A]

1. Prove Theorem 2.
2. Prove Theorem 3.
3. Prove Theorem 4.
4. Complete the proof of Corollary 3.
5. Prove Theorem 5.
6. If $P(A) = 0.6$, $P(B) = 0.5$, $P(A \cap B) = 0.3$, find $P(\overline{A})$, $P(\overline{B})$, and $P(A \cup B)$.
7. Why is it not possible to have $P(A) = 0.8$, $P(B) = 0.3$, $P(A \cap B) = 0.4$?
8. If C and D are mutually exclusive, what can be said about (a) $P(C \cap D)$? (b) $P(C \cup D)$?
9. If odds for E are 1 : 3, determine $P(E)$ and $P(\overline{E})$.
10. If $P(E) = 0.25$, find odds for E and against E.
11. Prove: If odds for E are $a : b$, then $\dfrac{a}{b} = \dfrac{P(E)}{P(\overline{E})}$.
12. If the 36 outcomes in the sample space for the two-dice experiment of Exercise 4, page 502, are equally likely, compute the probabilities of the following events.

    (a) $\{(1, 2), (2, 1)\}$ (b) $\{(x, y) : x + y = 3\}$

(c) $\{(x, y) : x + y = 4\}$
(d) $\{(x, y) : x + y \neq 4\}$
(e) $\{(x, y) : x + y < 7\}$
(f) $\{(x, y) : x = 2\}$
(g) $\{(x, y) : y \leq 3\}$
(h) $\{(x, y) : x = y\}$
(i) $\{(x, y) : x < y\}$
(j) $\{(x, y) : x + y = 1\}$
(k) $\{(x, y) : x + y = 7\} \cup \{(x, y) : x + y = 11\}$
(l) $\{(x, y) : x > 4\} \cap \{(x, y) : y < 3\}$
(m) $\{(x, y) : x > 4\} \cup \{(x, y) : y < 3\}$

13. A 2-card hand is dealt from a deck of cards. Determine the probabilities of the following events.
(a) a pair of aces
(b) a pair
(c) two spades
(d) no spades
(e) at least one ace
(f) at most one spade
(g) an ace and a spade
(h) an ace or a spade

14. A committee of 3 people is to be chosen by lot from the set of people $\{A, B, C, D, E\}$. Compute the probabilities of the following events.
(a) A is chosen.
(b) B is chosen.
(c) A and B are chosen.
(d) A or B is chosen.
(e) A is not chosen.
(f) A is chosen and B is not.
(g) A is chosen or B is not.
(h) Neither A nor B is chosen.

15.3 Generalized Probability

We have noted (page 503) that the classical definition of probability, in terms of equally likely outcomes, has serious deficiencies. It is useful for the analysis of simple games of chance, but even in these there is the possibility of confusion and disagreement. In the seventeenth century, for example, there was controversy among eminent mathematicians over so simple an experiment as tossing two coins. One opinion was that the outcomes in the sample space $\{HH, HT, TH, TT\}$ are equally likely, and that the probability of 2 heads (HH) is therefore $\frac{1}{4}$. Another opinion was that the outcomes in the sample space {0 heads, 1 head, 2 heads} are equally likely, and that the probability of 2 heads is therefore $\frac{1}{3}$. Modern opinion holds overwhelmingly that the probability of 2 heads is $\frac{1}{4}$. It is hard for most of us to believe that there could be serious dispute on this point, but there was. Even today, in problems more complicated than tossing two coins, there may be so many possible sample spaces that we can easily be confused about whether the outcomes in any one of them are all equally likely.

There are other difficulties in the "equally likely outcomes" approach. An important application of probability is to so-called vital statistics, to births and deaths, to the incidence of certain diseases. It is convenient to speak, for example, of the probability that a 30-year-old man will live for another 40 years. It is possible to make an intuitively acceptable estimate of this probability by using the proportion of 30-year-olds that live to be 70, but it is hard to see how this estimate can be explained in terms of equally likely outcomes. If a thumbtack is tossed in the air, it will land on its head or tipped over, and while it seems reasonable to associate a probability with each of these outcomes, it seems unreasonable to expect that they will be equally likely. If a bent coin lands heads 600 times in 1000 tosses, heads and tails do not seem to be equally likely, and we would like a theory of probability that would permit us to say that the probability of heads on a single toss of this coin is about 0.6. Indeed, it seems doubtful that the probability of heads on a toss of a coin fresh from the mint is exactly $\frac{1}{2}$; perhaps 0.51 or 0.498 would be more accurate.

For many reasons, then, the definition of probability in terms of equally likely outcomes has been found inadequate. Though the resulting theory applies adequately to a limited class of problems, it fails to fit a great variety of problems that seem essentially probabilistic in nature, such as those that arise in life insurance or in the toss of a bent coin. These considerations present the mathematician with a challenge: Find a definition of probability that will include all the useful features of the old definition but will also extend to problems that escape the classical theory. During the twentieth century, mathematicians have responded successfully to this challenge.

If a bent coin falls heads 600 times in 1000 tosses, it is reasonable to say that the probability of heads on a single toss is about 0.6. If out of 85,441 people aged 30 it was found that 38,569 lived to be 70, a reasonable estimate of the probability that a 30-year-old will live to be 70 is $\frac{38{,}569}{85{,}441} \approx 0.451$. If an average of only one out of 10,000 people gets some rare disease, it is reasonable to say that the probability that a randomly chosen person will get the disease is 0.0001. If 99 out of 100 students in a certain high school graduate, it is reasonable to say that a randomly chosen student has probability 0.99 of graduating.

Examples like these suggest that, in developing a modern theory of probability, it is desirable to permit great latitude in assigning probabilities to the possible outcomes of an experiment. The classical theory measures on a scale running from 0 to 1, as we proved in Section 15.2. As we shall see, a modern theory will be consistent with the classical theory in this respect, and yet permit the desired freedom in assigning probabilities, if, for any sample space S and any event $A \subseteq S$, we require only that (1) $P(A) \geq 0$, and (2) $P(S) = 1$.

The classical theory of probability derives from the properties of the counting function certain basic relationships that tell us how to combine probabilities. One of these key relationships is Theorem 1, page 504, concerning the probability of the union of two mutually exclusive events. If we replace the classical definition of probability, that is, $P(A) = \frac{n(A)}{n(S)}$, we no longer have a simple relationship between counting and probabilities, and we must therefore make some assumption about the probability of a union. The standard way to do this is to *assume* the validity of Theorem 1, which then becomes an axiom in our modern theory. These considerations lead to the following definition of the probability of an event.

▶ **Definition.** Given a sample space S, the probability of an event $A \subseteq S$ is a number $P(A)$ such that:

(1) For each $A \subseteq S$, $P(A) \geq 0$.
(2) $P(S) = 1$.
(3) For each $A, B \subseteq S$, $A \cap B = \emptyset \longrightarrow P(A \cup B) = P(A) + P(B)$.

It follows from the definition of a measure (page 476) that the probability function P is a measure with the special property that $P(S) = 1$.

We now discard the classical definition of probability (equation (*1*)), stated in terms of equally likely outcomes, and replace it with this new definition. In so doing, we discard the proofs of the theorems in Section 15.2, since those proofs depended on the classical definition. The new definition was designed, however, to include the classical theory as a special case. The theorems of the classical theory should therefore still be valid. What we must now do is to supply new proofs which are independent of (*1*). We list the theorems and corollaries here.

▶ **Theorem 1.** $A \cap B = \emptyset \longrightarrow P(A \cup B) = P(A) + P(B)$.

This has now become property (3) of our definition, and is therefore valid in the new theory.

▶ **Corollary 1.** $P(\emptyset) = 0$.

The proof in Section 15.2 depends only on Theorem 1. It is therefore still valid in the new theory.

▶ **Theorem 2.** Let $A_1, A_2, \ldots, A_k$ be events such that $i \neq j \longrightarrow A_i \cap A_j = \emptyset$. Then $P(A_1 \cup A_2 \cup \cdots \cup A_k) = \sum_{i=1}^{k} P(A_i)$.

The proof outlined in Section 15.2 depends on properties of the counting function and is therefore no longer valid. Since, however, the probability function P is a measure, this theorem is a special case of Exercise 2(**d**), page 484. Hence Theorem 2 holds in our new theory.

▶ **Theorem 3.** $P(\overline{A}) = 1 - P(A)$.

The old proof is no longer valid. It is left to the student as an exercise to supply a new proof, using properties (2) and (3) of our definition.

Corollary 2. $P(S) = 1$.

This is property (2) of our definition.

▶ **Theorem 4.** $A \subseteq B \longrightarrow P(A) \leq P(B)$.

The former proof is no longer valid. Since, however, P is a measure, this theorem is a special case of Exercise 2(**b**), page 484. Hence Theorem 4 holds in our new theory.

Corollary 3. For any event A, $0 \leq P(A) \leq 1$.

The proof indicated in Section 15.2 uses only Corollaries 1 and 2 and Theorem 4. Since these are valid in our new theory, the old proof is still good.

▶ **Theorem 5.** $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

This follows from Exercise 2(**c**), page 484.

We conclude that our modern theory of probability includes all the features of the classical theory that we proved in Section 15.2. We turn now to the details of applying the new theory. If we consider any one event A, the only necessary restriction on $P(A)$ is given by Corollary 3. If, for example, we take $\{H, T\}$ as sample space for the toss of the bent coin of page 510, then $P(\{H\})$ may be any number in the interval from 0 to 1. We may take $P(\{H\}) = 0.6$ if we wish; it then follows from Theorem 3 that $P(\{T\}) = 0.4$. We could also take $P(\{H\}) = P(\{T\}) = 0.5$; this amounts to assuming that both outcomes are equally likely. We could even take $P(\{H\}) = 1$, $P(\{T\}) = 0$, though this is inconsistent with getting 400 tails in 1000 throws.

Since the new definition permits so much latitude in fixing the probability of an event, it may be difficult to decide how to determine the probabilities in a particular case. In a great many cases, a procedure very like that of counting equally likely outcomes works well. Consider a sample space of N outcomes, $S = \{o_1, o_2, \cdots, o_N\}$. The subsets of this sample space that contain just one outcome each, namely $\{o_1\}$, $\{o_2\}$, $\cdots$, $\{o_N\}$, are called elementary events. The elementary events are mutually exclusive in pairs: $\{o_1\} \cap \{o_2\} = \emptyset$, $\{o_1\} \cap \{o_3\} = \emptyset$, and, in general, $i \neq j \longrightarrow \{o_i\} \cap \{o_j\} = \emptyset$. The union of all the elementary events is the whole sample space: $\{o_1\} \cup \{o_2\} \cup \cdots \cup \{o_N\} = S$. Hence, by Theorem 2, $P(S)$ is the sum of the probabilities of the elementary events, and, by property (2) of our definition, this sum is 1. The conditions of the definition will then be satisfied if we assign probabilities to the elementary events in such a way that (i) each such probability is non-negative, and (ii) their sum is 1. Any assignment of probabilities that satisfies these two conditions is acceptable, in that it is consistent with the mathematical theory. If we assign probability $\frac{1}{N}$ to each elementary event, which is the formal equivalent of the intuitive concept of equally likely outcomes, this is an acceptable assignment. In a great many cases, it is also a desirable assignment. It is not, however, the only acceptable assignment.

Every non-empty event $A \subseteq S$ is a union of elementary events. We see from Theorem 2 that we may then determine $P(A)$ by adding the probabilities of the elementary events whose union is A. Thus, once we have made an acceptable assignment of probabilities to the elementary events of S, the computation of $P(A)$ is just a matter of adding the appropriate probabilities. If we have made the uniform assignment $P(\{o_i\}) = \frac{1}{N}$ for all i, $1 \leq i \leq N$, and if the number of outcomes in A is m, then $P(A) = m \cdot \left(\frac{1}{N}\right) = \frac{m}{N}$. In this case, $P(A)$ is $\frac{n(A)}{n(S)}$, that is, the number of outcomes favorable to A divided by the total number of possible outcomes, and the modern theory is equivalent to the classical theory.

Example. A die is tossed. Find the probability that the number falling uppermost is (a) greater than 3, (b) odd.

Solution: There are infinitely many correct answers, depending on the probability assigned to each elementary event. We take as a sample space $S = \{1, 2, 3, 4, 5, 6\}$. If we make the uniform assignment of probability, $P(\{1\}) = P(\{2\}) = \cdots = P(\{6\}) = \frac{1}{6}$, we then have (a) $P(\{4, 5, 6\}) = P(\{4\}) + P(\{5\}) + P(\{6\}) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$, and (b) $P(\{1, 3, 5\}) = P(\{1\}) + P(\{3\}) + P(\{5\}) = \frac{1}{2}$. This is the natural assignment of probability if we believe that the die is perfectly symmetrical, an honest die. Suppose, however, that we suspect the die has been loaded in such a way that the probability of rolling a 1 is twice that of rolling any other number. We may then set $P(\{1\}) = \frac{2}{7}$, $P(\{2\}) = P(\{3\}) = \cdots = P(\{6\}) = \frac{1}{7}$, and relative to this assignment of probabilities we find **(a)** $P(\{4, 5, 6\}) = \frac{1}{7} + \frac{1}{7} + \frac{1}{7} = \frac{3}{7}$, and **(b)** $P(\{1, 3, 5\}) = \frac{2}{7} + \frac{1}{7} + \frac{1}{7} = \frac{4}{7}$. Still other assignments of probability are possible, and they will produce still other answers.

The student may be bewildered at this point, wondering how it is possible to produce correct solutions to problems when so many different assignments of probability are possible. This confusion is not peculiar to the subject of probability. In any application of mathematics to real problems of the physical world, the first step is an attempt to translate the conditions of the problem into mathematical terms. This step is often called *building a model* of the problem, and the success of any application of mathematics depends in a crucial way on the skill with which the model is constructed. To take an example from another branch of mathematics, by far the best-known geometrical model of the physical world is the Euclidean geometry studied in schools. For a long time it was believed that this was the only possible model, but in the nineteenth century mathematicians developed non-Euclidean geometries, and in the twentieth century physicists have shown that the Euclidean model is not satisfactory for the interpretation of certain physical events.

The model-building process in applications of probability consists in determining a sample space and assigning probabilities to the corresponding elementary events. The mathematical theory of probability does not tell us in detail how to do this. It tells us whether a sample space and an assignment of probabilities are *acceptable*, in a technical sense, but it cannot tell us whether they will be *useful* in interpreting random events. In the problem of the bent coin (page 510), common sense suggests that the assignment $P(\{H\}) = 0.6$, $P(\{T\}) = 0.4$ to the sample space $\{H, T\}$ will give us a realistic and useful model, though it may not be perfect. The assignment $P(\{H\}) = 0.1$, $P(\{T\}) = 0.9$ seems wildly unreasonable in the face of the evidence, 600 heads in 1000 tosses, and therefore unlikely to give us a useful model. Yet both models are technically acceptable, entirely consistent with the mathematical theory, and the theorems of probability apply equally well to both.

In most of the problems that follow, the student will find that there is an obvious, common-sense way to determine a sample space and assign probabilities. In particular, the "equally-likely outcomes" approach, in which equal probabilities are assigned to all elementary events, is by far the most common and is often suggested by certain special terminology. Thus, an *honest* (or *fair*, or *symmetrical*) coin is one for which $P(\{H\}) = P(\{T\}) = \frac{1}{2}$, and an *honest* die is one for which $P(\{1\}) = P(\{2\}) = \cdots = P(\{6\}) = \frac{1}{6}$. Similarly, if an r-element subset is drawn *at random* from a set of n elements, then each of the $\binom{n}{r}$ elementary events has probability $\binom{n}{r}^{-1}$.

Example 1. An honest coin is tossed three times. Find the probabilities of the following events.

(a) $A =$ "At least two heads." (b) $B =$ "Heads and tails alternate."
(c) $C =$ "All three tosses the same." (d) $A \cup C$

Solution: We take $\{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$ as a sample space. Since the coin is *honest*, we assign probability $\frac{1}{8}$ to each elementary event.

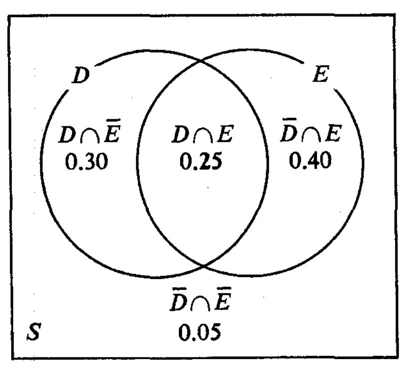
(a) $A = \{HHH, HHT, HTH, THH\} = \{HHH\} \cup \{HHT\} \cup \{HTH\} \cup \{THH\}$.

Hence, $$P(A) = \tfrac{1}{8} + \tfrac{1}{8} + \tfrac{1}{8} + \tfrac{1}{8} = \tfrac{1}{2}.$$

(b) $B = \{HTH, THT\}$. $P(B) = \frac{2}{8} = \frac{1}{4}$.
(c) $C = \{HHH, TTT\}$. $P(C) = \frac{2}{8} = \frac{1}{4}$.
(d) $A \cap C = \{HHH\}$. $P(A \cap C) = \frac{1}{8}$, and $P(A \cup C) = \frac{1}{2} + \frac{1}{4} - \frac{1}{8} = \frac{5}{8}$, by Theorem 5.

Example 2. $P(D) = 0.55$, $P(E) = 0.65$, $P(D \cap E) = 0.25$. Find (a) $P(\overline{D} \cap \overline{E})$, (b) $P(\overline{D} \cup E)$.

Solution: The figure is a Venn diagram showing events D and E in the sample space S. The four non-overlapping regions in the diagram correspond to four mutually exclusive events, as shown: $D \cap \overline{E}$, $D \cap E$, $\overline{D} \cap E$, $\overline{D} \cap \overline{E}$. We enter $P(D \cap E) = 0.25$ in the appropriate region. Because $P(D) = 0.55$, we must have $P(D \cap \overline{E}) = P(D) - P(D \cap E) = 0.55 - 0.25 = 0.30$ by property (3) of the definition; this information is also entered on the diagram. Similarly, we compute and enter $P(\overline{D} \cap E) = P(E) - P(D \cap E) = 0.65 - 0.25 = 0.40$.



(a) We have now found the probabilities of three of our four mutually exclusive events. Since the sum of the probabilities in the diagram must be $P(S) = 1$, we can find the probability of the fourth event: $P(\overline{D} \cap \overline{E}) = 1 - (0.30 + 0.25 + 0.40) = 1 - 0.95 = 0.05$. (b) $\overline{D} \cup E$ is represented in the diagram by those points that are not in the D circle or are in the E circle. We see that $\overline{D} \cup E$ is therefore the union of the events $D \cap E$, $\overline{D} \cap E$, and $\overline{D} \cap \overline{E}$. Since these events are mutually exclusive in pairs, it follows from Theorem 2 that $P(\overline{D} \cup E) = P(D \cap E) + P(\overline{D} \cap E) + P(\overline{D} \cap \overline{E}) = 0.25 + 0.40 + 0.05 = 0.70$.

Example 3. Two balls are drawn at random, without replacement, from an urn that contains 4 red balls and 3 white balls. Compute the probabilities of the following events.

(a) $F =$ "Both balls red."
(b) $G =$ "Both balls white."
(c) $H =$ "Both balls the same color."
(d) $I =$ "Balls are of different colors."

Solution: We take as a sample space the set of $\binom{7}{2} = 21$ possible pairs of balls, and, since the drawings are *at random*, we assign probability $\frac{1}{21}$ to each elementary event.

(a) The number of ways of choosing 2 red balls out of 4 is $\binom{4}{2} = 6$. Since we have equally likely outcomes, it follows that $P(F) = \frac{6}{21} = \frac{2}{7}$.

(b) There are $\binom{3}{2} = 3$ ways of choosing the 2 white balls. Hence, $P(G) = \frac{3}{21} = \frac{1}{7}$.

(c) H is the union of the mutually exclusive events F and G: $F \cap G = \emptyset$, $F \cup G = H$. Hence, $P(H) = P(F) + P(G) = \frac{2}{7} + \frac{1}{7} = \frac{3}{7}$.

(d) We can choose a red ball in $\binom{4}{1}$ ways, a white ball in $\binom{3}{1}$ ways. By the fundamental principle of counting, the number of ways of choosing a red and a white is then $\binom{4}{1}\binom{3}{1} = 12$, and therefore $P(I) = \frac{12}{21} = \frac{4}{7}$. Alternatively, $I = \overline{H}$, and therefore $P(I) = 1 - P(H) = 1 - \frac{3}{7} = \frac{4}{7}$.

Exercises [A]

1. Prove Theorem 3, page 512.

2. $S = \{o_1, o_2, o_3\}$ is a sample space. If $P(\{o_1\}) = P(\{o_2\}) = [P(\{o_3\})]^2$, find (a) $P(\{o_3\})$, (b) $P(\{o_1, o_2\})$.

3. (a) In the coin-tossing experiment of Example 1, page 515, find the probability of (i) 0 heads, (ii) 1 head, (iii) 2 heads, (iv) 3 heads.

   (b) (i) Is {0 heads, 1 head, 2 heads, 3 heads} an acceptable sample space for this experiment? (ii) Is the assignment of equal probabilities to the elementary events of this sample space acceptable? (iii) Does this assignment seem likely to produce a useful model for this experiment?

   (c) In the sample space of Example 1, page 515, let $E = \{TTT\}$. Describe the event $\overline{E}$ in words, and use Theorem 3 to compute $P(\overline{E})$.

(d) An honest coin is tossed 10 times. We take as a sample space the set S of all ordered 10-element sequences (ordered 10-tuples) of H's and T's. (i) How many elements are there in S? (ii) Compute P(At least one head).

4. Events $A, B \subseteq S$ are such that $P(A) = 0.5$, $P(B) = 0.6$, $P(A \cap B) = 0.4$. Draw a Venn diagram, enter the appropriate probability in each of the non-overlapping regions of the diagram, and determine the following: (a) $P(A \cup B)$, (b) $P(\bar{A})$, (c) $P(\bar{B})$, (d) $P(\bar{A} \cap \bar{B})$, (e) $P(\bar{A} \cup \bar{B})$, (f) $P(A \cap \bar{B})$, (g) $P(\bar{A} \cap B)$, (h) $P(\bar{A} \cup B)$.

5. Show that it is not possible to have $P(C) = 0.8$, $P(D) = 0.6$, $P(C \cap D) = 0.3$.

6. Out of 200 high-school graduates, it was found that 83 had studied chemistry, 67 had studied physics, and 20 had studied both sciences. What is the probability that a randomly selected member of this set had studied neither chemistry nor physics?

7. A card is drawn at random from a full pack. What is the probability that it is (a) an ace or a spade? (b) an honor or a spade? (The honor cards are the 10, J, Q, K, A of each suit.) (c) an honor or not a spade?

8. A card is drawn at random from a deck. It is replaced, and a second random drawing is made.

   (a) Describe a sample space for this experiment, and assign probabilities to the elementary events.

   (b) Compute the probability that the first card drawn is a spade or the second is a spade.

9. A subcommittee of 3 members is to be chosen by lot from a committee of 4 Democrats and 3 Republicans. What is the probability that

   (a) the subcommittee is all Democrats?

   (b) both parties are represented?

10. Two letters are chosen at random from $P\,A\,R\,R\,A\,M\,A\,T\,T\,A$. What is the probability that

    (a) they are the same? (b) they are different

    (c) one of them is P? (d) one of them is A?

11. An urn contains 3 black balls, 2 red balls, and 4 white balls. Two are withdrawn at random. What is the probability that they are of different colors?

12. A die is loaded in such a way that, in the long run, 2, 3, 4, and 5 each come up about twice as often as 1, and 6 comes up about three times as often as 1.
    (a) Determine a sample space for the experiment of rolling this die once, and assign probabilities to the elementary events.
    (b) What is the probability of rolling an odd number?
    (c) What are the odds against rolling a 2?

13. A die is loaded in such a way that the probability that any face turns up is proportional to the number on that face.
    (a) Determine a sample space for the experiment of rolling this die once, and assign probabilities to the elementary events.
    (b) What is the probability of rolling a number greater than 3?

14. Use a Venn diagram to show that, for any events A, B, $C \subseteq S$, $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - P(C \cap A) + P(A \cap B \cap C)$.

15. If a card is drawn at random from a full pack, what is the probability that it is (a) black, a spade, or an ace? (b) red, a spade, or an honor? (c) a spade, an ace, or an honor?

16. The kings and queens of hearts and diamonds are put in one pile, the kings and queens of spades and clubs in another. Two cards are drawn, without replacement. What is the probability that both cards are kings if
    (a) both are taken from one pile? (b) one is taken from each pile?
    (c) the piles are first mixed together, and then both are taken from the combined pile?

15.4 Independent Events

If we take a deck of cards, draw a card at random, and then make a second drawing without replacing the card drawn on the first, it is clear that the outcome of the second drawing depends on the outcome of the first. If, for example, we get the ace of spades on the first drawing, it is then impossible to get the ace of spades again on the second drawing. On the other hand, if we return the first card to the deck and shuffle thoroughly before making the second drawing, common sense suggests that the outcome of the second drawing is in no way affected by the outcome of the first: the two drawings are *independent.* It will be our concern in this section to make explicit this intuitive concept of independence and to explore its use as a mathematical tool.

Consider further the experiment of two drawings, with replacement, from a deck of cards. Since there are 52 cards in the deck, the sample space of all possible ordered pairs of drawings contains $(52)^2 = 2704$ points, and the assumption of randomness leads us to assign probability $\frac{1}{2704}$ to each of the corresponding elementary events. Now let $E =$ "First draw a spade." Since there are 13 spades, and since for each spade on the first draw there are 52 possibilities for the second draw, there are $13 \cdot 52 = 676$ points in E, and therefore $P(E) = \frac{676}{2704} = \frac{1}{4}$. Let $F =$ "Second draw a face card" (the face cards are the king, queen, and jack of each suit). Since there are $4 \cdot 3 = 12$ face cards, and since each face card on the second draw may be paired with any one of 52 cards on the first draw, there are $52 \cdot 12 = 624$ points in F, and therefore $P(F) = \frac{624}{2704} = \frac{3}{13}$. Finally, we consider $E \cap F =$ "First draw a spade and second draw a face card." There are $13 \cdot 12 = 156$ points in $E \cap F$, and therefore $P(E \cap F) = \frac{156}{2704} = \frac{3}{52}$. We note that, in this case,

$$P(E \cap F) = P(E)P(F).$$

Suppose now that we draw one card at random from the deck. Let $E =$ "Spade" and $F =$ "Face card." Because of the symmetry of the deck (each suit contains the same number of face cards), common sense again suggests that E and F are independent; whether a card is a face card has nothing to do with whether it is a spade. Since there are 13 spades and 12 face cards, we find $P(E) = \frac{13}{52} = \frac{1}{4}$, $P(F) = \frac{12}{52} = \frac{3}{13}$. Since the spade suit contains 3 face cards, $P(E \cap F) = \frac{3}{52}$. Again,

$$P(E \cap F) = P(E)P(F).$$

By contrast, let us examine a case in which events are intuitively not independent. In the experiment of tossing a fair coin three times, let $E =$ "First toss heads" and $F =$ "At least two heads." It seems evident that getting heads on the first toss should increase the probability of getting at least two heads in three tosses, and common sense therefore suggests that E and F are not independent. We compute

$$\begin{aligned} P(E) &= P(\{HHH, HHT, HTH, HTT\}) = \tfrac{1}{2}, \\ P(F) &= P(\{HHH, HHT, HTH, THH\}) = \tfrac{1}{2}, \\ P(E \cap F) &= P(\{HHH, HHT, HTH\}) = \tfrac{3}{8}. \end{aligned}$$

We observe that, in this case, $P(E \cap F) \neq P(E)P(F)$.

The intuitive concept of independent events as events that, in some sense, do not affect one another, or have nothing to do with one another, is too vague to use as a mathematical definition. We have seen, however, that in at least two cases the intuitive notion of independence is associated with the relationship $P(E \cap F) = P(E)P(F)$, and that in at least one case this relationship was not satisfied when events were intuitively not independent. Since these observations are typical, we take this relationship as our definition of independent events.

▶ **Definition.** *Events $E, F \subseteq S$ are independent if and only if*

$$P(E \cap F) = P(E)P(F). \tag{2}$$

If E and F are not independent, then they are dependent.

A word of caution is now in order. We have established a definition, and to determine whether or not two events are independent is no longer an intuitive matter. We determine a sample space and assign probabilities. If the multiplication rule (*2*) holds, then E and F are independent; if it does not hold, then they are dependent, no matter what intuition or common sense may urge. Generally, we shall find that (*2*) does what we naturally expect of it. If this were not so, we would replace it with a better definition.

Example 1. An honest coin and an honest die are tossed. Let $E =$ "Coin shows heads" and $F =$ "Die shows 5." Are E and F independent?

Solution: We take as sample space the set of $2 \cdot 6 = 12$ ordered pairs (outcome with coin, outcome with die), and assign probability $\frac{1}{12}$ to each elementary event. Then $P(E) = P(\{(H, 1), (H, 2), \ldots, (H, 6)\}) = \frac{6}{12} = \frac{1}{2}$, $P(F) = P(\{(H, 5), (T, 5)\}) = \frac{2}{12} = \frac{1}{6}$, and $P(E \cap F) = P(\{(H, 5)\}) = \frac{1}{12}$. Since $(\frac{1}{2})(\frac{1}{6}) = \frac{1}{12}$, E and F are independent.

Example 2. A fair coin is tossed twice. Let $C =$ "At least one head," $D =$ "At least one tail." Are C and D independent?

Solution: $P(C) = P(\{HH, HT, TH\}) = \frac{3}{4}$, $P(D) = P(\{HT, TH, TT\}) = \frac{3}{4}$, $P(C \cap D) = P(\{HT, TH\}) = \frac{1}{2}$. Since $(\frac{3}{4})(\frac{3}{4}) = \frac{9}{16} \neq \frac{1}{2}$, events C and D are dependent.

Example 3. An honest coin is tossed twice. Let $E =$ "Not more than one head," $F =$ "All tosses alike." Are E and F independent?

Solution: $P(E) = P(\{HT, TH, TT\}) = \frac{3}{4}$, $P(F) = P(\{HH, TT\}) = \frac{1}{2}$, $P(E \cap F) = P(\{TT\}) = \frac{1}{4}$. Since $(\frac{3}{4})(\frac{1}{2}) = \frac{3}{8} \neq \frac{1}{4}$, E and F are dependent.

Example 4. Modify Example 3 by tossing three times.

Solution: $P(E) = P(\{HTT, THT, TTH, TTT\}) = \frac{4}{8} = \frac{1}{2}$, $P(F) = P(\{HHH, TTT\}) = \frac{2}{8} = \frac{1}{4}$, $P(E \cap F) = P(\{TTT\}) = \frac{1}{8}$. Since $(\frac{1}{2})(\frac{1}{4}) = \frac{1}{8}$, E and F are independent in this case.

Once we have assigned probabilities to the elementary events of a sample space, the question of independence is settled by our definition: events E

and F are independent if and only if their probabilities satisfy the multiplication rule (*2*). On the other hand, if we have an intuitive notion that two events ought to be independent, we can often make our probabilistic model conform to this judgment by using (*2*) in assigning probabilities. This procedure is particularly appropriate in compound experiments, such as repeated tossing of a coin or a die, in which the successive operations are intuitively independent.

Example. A bent coin has probability $\frac{2}{3}$ of showing heads on any one toss. It is tossed twice. Use (*2*) to assign probabilities to the elementary events of $S = \{HH, HT, TH, TT\}$.

Solution: Because the outcomes of the two tosses are intuitively independent, we set $P(\{HH\}) = P((\text{Heads on first toss}) \cap (\text{Heads on second toss})) = P(\text{Heads on first toss})P(\text{Heads on second toss}) = (\frac{2}{3})(\frac{2}{3}) = \frac{4}{9}$. Similarly, $P(\{HT\}) = P((\text{Heads on first toss}) \cap (\text{Tails on second toss})) = P(\text{Heads on first toss})P(\text{Tails on second toss}) = (\frac{2}{3})(\frac{1}{3}) = \frac{2}{9}$, $P(\{TH\}) = (\frac{1}{3})(\frac{2}{3}) = \frac{2}{9}$, $P(\{TT\}) = (\frac{1}{3})(\frac{1}{3}) = \frac{1}{9}$.

The concept of independence can be extended to three or more events. We say that such events are independent if the multiplication rule (*2*) or an extension of (*2*) holds for *every intersection* of two or more of the events. Thus, for example, events E_1, E_2, E_3 are independent if and only if

$$\begin{aligned} P(E_1 \cap E_2) &= P(E_1)P(E_2) \\ P(E_2 \cap E_3) &= P(E_2)P(E_3) \\ P(E_3 \cap E_1) &= P(E_3)P(E_1) \\ P(E_1 \cap E_2 \cap E_3) &= P(E_1)P(E_2)P(E_3). \end{aligned}$$

One might expect that the first three of these equations would imply the fourth, but such is not the case.

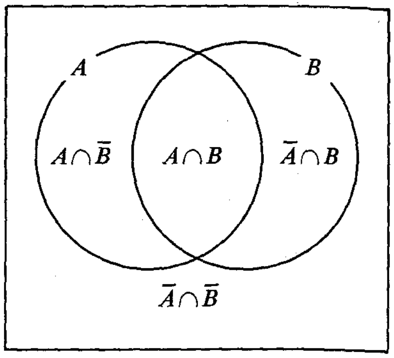
Example. A red and a green die are tossed. Let E_1 = "Red die shows odd number," E_2 = "Green die shows odd number," E_3 = "Sum of numbers is odd." Are these events independent?

Solution: Referring to the sample space and assignment of probabilities that we made for Exercise 4, page 502, we find $P(E_1) = P(E_2) = P(E_3) = \frac{18}{36} = \frac{1}{2}$ and $P(E_1 \cap E_2) = P(E_2 \cap E_3) = P(E_3 \cap E_1) = \frac{9}{36} = \frac{1}{4}$. Since $\frac{1}{4} = (\frac{1}{2})(\frac{1}{2})$, it follows that E_1, E_2, E_3 are independent in pairs. But $P(E_1 \cap E_2 \cap E_3) = 0 \neq (\frac{1}{2})(\frac{1}{2})(\frac{1}{2})$, and the events are therefore not independent.

We conclude this consideration of independence with a useful theorem.

▶ **Theorem 6.** If A and B are independent events, then so are **(a)** A and $\bar{B}$, **(b)** $\bar{A}$ and B, **(c)** $\bar{A}$ and $\bar{B}$.

Proof: We prove only (a), leaving (b) and (c) as exercises. We note from the figure to the right that $A \cap \overline{B}$ and $A \cap B$ are mutually exclusive events, and that their union is A. It follows from the definition of probability that



$$P(A) = P(A \cap \overline{B}) + P(A \cap B). \quad (3)$$

But A and B are given as independent, and therefore $P(A \cap B) = P(A)P(B)$. Hence (3) becomes

$$P(A) = P(A \cap \overline{B}) + P(A)P(B),$$

or

$$\begin{aligned} P(A \cap \overline{B}) &= P(A) - P(A)P(B) \\ &= P(A)[1 - P(B)] \\ &= P(A)P(\overline{B}). \qquad \text{(by Theorem 3)} \end{aligned}$$

Thus A and $\overline{B}$ are independent by definition.

In working with the intersections of two events and their complements, it is often convenient to display the probabilities involved in a table.

	A	$\overline{A}$	
B	$P(A \cap B)$	$P(\overline{A} \cap B)$	$P(B)$
$\overline{B}$	$P(A \cap \overline{B})$	$P(\overline{A} \cap \overline{B})$	$P(\overline{B})$
	$P(A)$	$P(\overline{A})$	1

The body of this table is simply an alternate form of the Venn diagram above. There are four cells in the body of the table, one corresponding to each of the four regions in the Venn diagram. The corresponding probability is entered in each cell. By (3), the sum of the two entries in the first column of the table is $P(A)$; this is shown in the margin of the table, at the bottom. Similarly, the sum of the second column is $P(\overline{A})$, also entered at the bottom. The sums of the two rows are $P(B)$ and $P(\overline{B})$, entered in the right-hand margin. The sum of the marginal row is 1, by Theorem 3, as is the sum of the marginal column.

Example 1. A red and a green die are rolled. Let A = "Red die shows 2," B = "Sum of numbers is 5." Construct a table showing the probability of each intersection of A or $\overline{A}$ with B or $\overline{B}$.

Solution: We assign probability $\frac{1}{36}$ to each elementary event in the sample space of Exercise 4, page 502. Then $P(A \cap B) = P(\{(2, 3)\}) = \frac{1}{36}$, and so on.

	A	$\overline{A}$	
B	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{4}{36}$
$\overline{B}$	$\frac{5}{36}$	$\frac{27}{36}$	$\frac{32}{36}$
	$\frac{6}{36}$	$\frac{30}{36}$	1

If A and B are independent, then Theorem 6 tells us that each entry in the body of a table like this is the product of the corresponding marginal entries. The table then takes the form of a multiplication table.

Example 2. Repeat the above example with $B =$ "Green die shows 3."

Solution:

Each entry in the body of this table is the product of the corresponding marginal entries. In particular, $\frac{1}{36} = (\frac{6}{36})(\frac{6}{36})$, or $P(A \cap B) = P(A)P(B)$. Hence A and B are independent. Compare this table with the table of Example 1 above, in which A and B are dependent.

	A	$\overline{A}$	
B	$\frac{1}{36}$	$\frac{5}{36}$	$\frac{6}{36}$
$\overline{B}$	$\frac{5}{36}$	$\frac{25}{36}$	$\frac{30}{36}$
	$\frac{6}{36}$	$\frac{30}{36}$	1

Exercises [A]

1. Prove for any event $A \subseteq S$.
   (a) A and $\emptyset$ are independent.
   (b) A and S are independent.
2. What is $P(B)$ if B and B are independent?
3. E and F are independent events. The probability that both occur is $\frac{1}{24}$, and the probability that neither occurs in $\frac{7}{12}$. Find $P(E)$ and $P(F)$.
4. An honest coin is tossed five times. What is the probability that heads and tails alternate?
5. An honest die is rolled four times. Find the probability of a 3 on the first roll, a number greater than 3 on the second, a number less than 3 on the third, and 3 again on the fourth roll.
6. (a) Determine a sample space and assign the probabilities to the elementary events if the bent coin of the example on page 521 is tossed three times.
   (b) What is the probability of getting exactly two heads in the three tosses?
   (c) Are the events of Examples 3 and 4 on page 520 independent with this assignment of probabilities?

7. In Exercise 6, page 517, test for independence the events "Student selected has studied chemistry" and "Student selected has studied physics."

8. A card is drawn from a deck. Let A = "Card is an ace," B = "Card is a spade," C = "Card is an honor."
   (a) Make a table showing the probability of each intersection of A or $\overline{A}$ with B or $\overline{B}$. Are A and B independent?
   (b) Repeat, using B and C.

9. Prove parts (b) and (c) of Theorem 6.

10. How many times must a fair coin be tossed so that $P(\text{No heads}) \leq 0.01$?

11. Four fair dice are thrown. If the outcomes on the dice are independent of one another, find the probability of each of the following events.
    (a) Exactly three faces are alike. (b) At least one 2 is obtained.
    (c) The sum of the numbers is 6.

12. If the probability is 0.01 that a man crossing a busy highway will be hit by a car, and a foolish man habitually crosses twice a day, what is the probability that he will cross without accident for 30 days? (Use logarithms or a slide rule to evaluate.)

13. If the probability is 0.9928 that a 20-year-old will live to be 25, compute the probability of each of the following events.
    (a) Two 20-year-olds will both live to be 25.
    (b) Exactly one of two 20-year-olds will live to be 25.
    (c) Neither of two 20-year-olds will live to be 25.

15.5 Conditional Probability

An urn contains one red ball and two white balls. An experiment consists of drawing the three balls from the urn, one at a time. We ask for the probability of the event A = "Second ball drawn is red." If we note the colors of the balls as they are drawn (R for red, W for white), then a sample space for this experiment is $S = \{RWW, WRW, WWR\}$. Since the drawing is at random, it is reasonable to assign equal probabilities to the three elementary events in this sample space. We therefore conclude that $P(A) = \frac{1}{3}$.

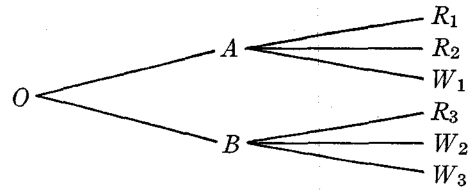
Suppose, now, that we modify this experiment by looking at the first ball drawn before we determine the probability of the event A. To take a specific instance, suppose that the first ball drawn is white. What then is the probability that the second ball will be red? Evidently, since one white ball has been removed from the urn, there are two balls left, one of them red. The probability of a red ball on the second draw is therefore $\frac{1}{2}$. Thus, if $B =$ "First ball drawn is white," then the probability of A, given that B has occurred, is $\frac{1}{2}$. This new probability is called the conditional probability of A given B, and is written $P(A \mid B)$. The original sample space for the computation of $P(A)$ was $\{RWW, WRW, WWR\}$. The fact that B has occurred in effect eliminates the outcome RWW from the sample space. It leaves us $B = \{WRW, WWR\}$ as a reduced sample space for the computation of the probability of A. The probability of A in this reduced sample space B is the conditional probability $P(A \mid B)$. It is not the same as the probability $P(A)$ that A had in the original sample space S. The additional information, the knowledge that B has occurred, changes our analysis of the experiment and therefore changes our computed probability of A.

Similarly, let $C =$ "First ball drawn is red," and draw one ball from the urn. If C occurs, then our sample space for the computation of the probability of A is reduced to just $\{RWW\}$. Since A cannot occur in this sample space, we conclude that $P(A \mid C) = 0$.

The general situation is this. The probability of an event A is the probability that the outcome o of an experiment belongs to the set A: $P(A) = P(o \in A)$. We are told that an event B has occurred, that is, that $o \in B$. In computing the resulting conditional probability, we then take B as a reduced sample space for the experiment and compute the new probability of A, $P(A \mid B)$, in this reduced sample space. Note that since we are given $o \in B$, we conclude that if A also occurs, that is, if $o \in A$, then $o \in A \cap B$.

Example 1. Urn A contains two red balls and one white ball, urn B contains one red ball and two white balls. An urn is chosen at random, and a ball is withdrawn from that urn. Let $E =$ "Ball chosen is white," $F =$ "Urn A is chosen." Compute $P(E \mid F)$.

Solution: The figure shows a tree for this experiment. The balls are numbered to help us keep track of all possible outcomes. We see from the tree that there are 6 outcomes in our sample space S, and using the assumption of randomness and

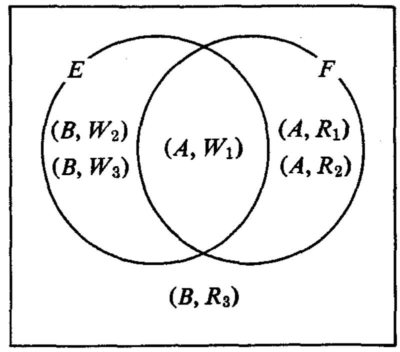


Outcome
(A, R_1)
(A, R_2)
(A, W_1)
(B, R_3)
(B, W_2)
(B, W_3)

the symmetry of the tree, we assign probability $\frac{1}{6}$ to each elementary event. Since $E = \{(A, W_1), (B, W_2), (B, W_3)\}$, $F = \{(A, R_1), (A, R_2), (A, W_1)\}$, we have $P(E)=\frac{3}{6}=\frac{1}{2}=P(F)$. The event F serves as a reduced sample space for the computation of $P(E \mid F)$. There are 3 equally likely outcomes in F. Just one of these, namely (A, W_1), is also in E. Hence, $P(E \mid F)=\frac{1}{3}$.

A Venn diagram may clarify the preceding argument. There are 6 sample points in S, 3 of which are in E, and, since all outcomes in S are equally likely, it follows that $P(E)=\frac{3}{6}=\frac{1}{2}$. If, however, we are given that F has occurred, we are then restricted to the reduced sample space F. There are just 3 sample points in F, one of which is also in E (that is, in $E \cap F$), and therefore $P(E \mid F)=\frac{1}{3}$. Note that, in this case,

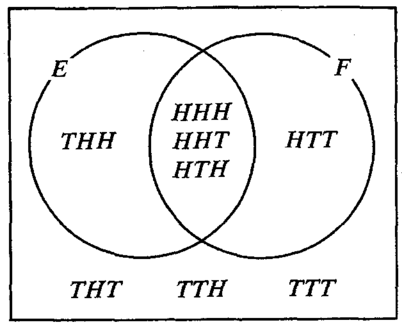
$$P(E \mid F)=\frac{\frac{1}{6}}{\frac{3}{6}}=\frac{P(E \cap F)}{P(F)}.$$



Example 2. A fair coin is tossed three times. Find the probability of getting at least two heads, given that the first toss is heads.

Solution: We take $S=\{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$, $E=$ "At least two heads" $=\{HHH, HHT, HTH, THH\}$, $F=$ "First toss heads" $=\{HHH, HHT, HTH, HTT\}$. We assign probability $\frac{1}{8}$ to each elementary event. Then $P(E)=\frac{4}{8}=\frac{1}{2}=P(F)$. In the reduced sample space F (see the Venn diagram to the right), there are 4 sample points, 3 of which are also in E (and therefore in $E \cap F$). Hence, $P(E \mid F)=\frac{3}{4}$. Note that in this example we again have

$$P(E \mid F)=\frac{\frac{3}{8}}{\frac{4}{8}}=\frac{P(E \cap F)}{P(F)}.$$



In the two preceding examples, we had a uniform assignment of equal probabilities to all elementary events. The computation of probabilities was thereby reduced to counting sample points. We now consider a somewhat more sophisticated example.

Example 3. The bent coin of the example on page 521, which has probability $\frac{2}{3}$ of landing heads on any one toss, is tossed three times. Find the probability of getting at least two heads, given that the first toss is heads.

Solution: The sample space S and the events E and F are as defined in Example 2, but the assignment of probabilities is a little more complicated (see Exercise 6, page 523). We assume that the outcomes of successive tosses are independent. Hence, $P(\{HHH\}) = (\frac{2}{3})(\frac{2}{3})(\frac{2}{3}) = \frac{8}{27}$, $P(\{HHT\}) = (\frac{2}{3})(\frac{2}{3})(\frac{1}{3}) = \frac{4}{27} = P(\{HTH\}) = P(\{THH\})$, $P(\{HTT\}) = (\frac{2}{3})(\frac{1}{3})(\frac{1}{3}) = \frac{2}{27} = P(\{THT\}) = P(\{TTH\})$, $P(\{TTT\}) = (\frac{1}{3})(\frac{1}{3})(\frac{1}{3}) = \frac{1}{27}$. We may now easily compute $P(E) = \frac{8}{27} + \frac{4}{27} + \frac{4}{27} + \frac{4}{27} = \frac{20}{27}$, $P(F) = \frac{8}{27} + \frac{4}{27} + \frac{4}{27} + \frac{2}{27} = \frac{18}{27}$. Since the event F has occurred, we use F as a reduced sample space for the computation of $P(E \mid F)$. Examining the Venn diagram above, we see that there are 3 sample points in $E \cap F$: HHH, HHT, HTH. The probability associated with these points (in the sample space S) is $P(E \cap F) = \frac{8}{27} + \frac{4}{27} + \frac{4}{27} = \frac{16}{27}$. Since $P(F) = \frac{18}{27}$, it follows that $P(E \cap F)$ is just $\frac{16}{18}$ of the total probability of F, and it is natural to take this ratio as the probability of E given F: $P(E \mid F) = \frac{16}{18} = \frac{8}{9}$.

In Example 3 we were led by intuitive considerations to take the ratio $\dfrac{P(E \cap F)}{P(F)}$ as $P(E \mid F)$. Since the results obtained by counting sample points in Examples 1 and 2 can also be expressed as the ratio $\dfrac{P(E \cap F)}{P(F)}$, we now formalize this procedure.

▶ **Definition.** **If $E, F \subseteq S$, $P(F) > 0$, then the conditional probability of E given F is**

$$P(E \mid F) = \frac{P(E \cap F)}{P(F)}. \tag{4}$$

$P(E \mid F)$ is not defined if $P(F) = 0$.

Example 4. Two cards are dealt from a standard deck. If one of them is an ace, what is the probability that both of them are aces?

Solution: Let $E =$ "Both cards aces," $F =$ "At least one card an ace." Now, there are 4 aces in the deck, so that $\binom{4}{2}$ of the $\binom{52}{2}$ possible 2-card hands contain 2 aces. Since $E \subseteq F$, we have $E \cap F = E$, and therefore $P(E \cap F) = P(E) = \binom{4}{2} \Big/ \binom{52}{2} = \frac{1}{221}$. In finding $P(F)$, it is convenient to compute $P(\overline{F})$ and use Theorem 3. Since $\overline{F} =$ "No aces," and there are 48 non-aces in the pack, we have $P(\overline{F}) = \binom{48}{2} \Big/ \binom{52}{2} = \frac{188}{221}$, and therefore $P(F) = 1 - P(\overline{F}) = \frac{33}{221}$. Hence, $P(E \mid F) = \dfrac{\frac{1}{221}}{\frac{33}{221}} = \frac{1}{33}$.

The following theorem states a useful property of conditional probability.

▶ **Theorem 7.** **In the reduced sample space F, the conditional probability $P(E \mid F)$ satisfies the definition of the probability of an event.**

Proof: We must show that $P(E \mid F)$ has the three properties listed in the definition, page 511.

(1) (We must prove that $P(E \mid F) \geq 0$.)
Since $P(E \cap F) \geq 0$ by definition of the probability of an event, and since, in this case, $P(F) > 0$ by definition of conditional probability, we have

$$P(E \mid F) = \frac{P(E \cap F)}{P(F)} \geq 0.$$

(2) (Since the reduced sample space F here replaces S, we must prove that $P(F \mid F) = 1$.)

$$\begin{aligned} P(F \mid F) &= \frac{P(F \cap F)}{P(F)} && \text{(Definition of conditional probability)} \\ &= \frac{P(F)}{P(F)} && (F \cap F = F) \\ &= 1. \end{aligned}$$

(3) ($A \cap B = \emptyset \longrightarrow P((A \cup B) \mid F) = P(A \mid F) + P(B \mid F)$.)

We need two of the distributive properties listed in Exercise 11, page 483. First, we prove that if A and B are mutually exclusive, as given here, then also $A \cap F$ and $B \cap F$ are mutually exclusive:

$$\begin{aligned} (A \cap F) \cap (B \cap F) &= (A \cap B) \cap F && \text{(Ex. 11(d), page 484)} \\ &= \emptyset \cap F && \text{(Given, } A \cap B = \emptyset) \\ &= \emptyset. \end{aligned}$$

Second, by Exercise 11(**a**), page 483, we have:

$$(A \cap F) \cup (B \cap F) = (A \cup B) \cap F.$$

It follows, by part (3) of the definition of the probability of an event, that

$$\begin{aligned} P((A \cup B) \cap F) &= P((A \cap F) \cup (B \cap F)) \\ &= P(A \cap F) + P(B \cap F). \end{aligned}$$

We now have:

$$\begin{aligned} P((A \cup B) \mid F) &= \frac{P((A \cup B) \cap F)}{P(F)} && \text{(Definition of conditional probability)} \\ &= \frac{P(A \cap F) + P(B \cap F)}{P(F)} && \text{(by the preceding result)} \\ &= \frac{P(A \cap F)}{P(F)} + \frac{P(B \cap F)}{P(F)} \\ &= P(A \mid F) + P(B \mid F) && \text{(Definition of conditional probability)} \end{aligned}$$

Another intuitively desirable property of conditional probability is the following: If E does not depend on F, that is, if E and F are independent, then the fact that F has occurred should not affect the probability that E will occur. This is indeed the case.

▶ **Theorem 8.** If E and F are independent, then $P(E \mid F) = P(E)$.

Proof: If E and F are independent, then $P(E \cap F) = P(E)P(F)$. In these circumstances, we have

$$\begin{aligned} P(E \mid F) &= \frac{P(E \cap F)}{P(F)} \quad \text{(Definition of conditional probability)} \\ &= \frac{P(E)P(F)}{P(F)} \quad \text{(Definition of independence)} \\ &= P(E). \end{aligned}$$

If we clear (*4*) of fractions, we obtain

$$P(E \cap F) = P(F)P(E \mid F). \qquad (5)$$

This formula is often useful for computing $P(E \cap F)$, especially in compound experiments in which the successive operations are not independent. Note that (*5*) is a generalization of the definition of independent events. If E and F happen to be independent and $P(F) > 0$), then equation (*5*) reduces to equation (*2*) with the help of Theorem 8.

Example 1. Three defective light bulbs get mixed in with six good ones, and they are tested, one by one, until all the bad bulbs have been located. What is the probability that the seventh bulb tested is the third defective?

Solution: Let $F =$ "Two defectives among first six bulbs tested" and $E =$ "Seventh bulb tested is defective." The required probability is then $P(E \cap F)$, and by (*5*) this is equal to $P(F)P(E \mid F)$. To find $P(F)$, we note that there are $\binom{9}{6}$ ways to choose the first 6 bulbs tested. Since there are $\binom{3}{2}$ ways of choosing 2 defective bulbs and $\binom{6}{4}$ ways of choosing 4 good bulbs, then there are $\binom{3}{2}\binom{6}{4}$ ways of locating 2 defectives among the first 6 bulbs tested, by the fundamental principle of counting. Hence, assuming randomness, $P(F) = \binom{3}{2}\binom{6}{4} \Big/ \binom{9}{6} = \frac{15}{28}$. If F has occurred, then there is just 1 defective bulb among the 3 remaining, and the probability that it is the next one tested is then $P(E \mid F) = \frac{1}{3}$. Thus, $P(E \cap F) = \left(\frac{15}{28}\right)\left(\frac{1}{3}\right) = \frac{5}{28}$.

Example 2. Urn A contains 1 red, 3 white, and 4 blue balls. Urn B contains 2 red, 1 white, and 3 blue balls. An urn is chosen at random, and a ball is drawn from it. What is the probability that the ball is white?

Solution: Let $E =$ "Ball is white." It is both natural and necessary to consider separately the mutually exclusive events $F_1 =$ "Urn A is chosen" and $F_2 =$ "Urn B is chosen." We have proved that if F_1 and F_2 are mutually exclusive, then $E \cap F_1$ and $E \cap F_2$ are also mutually exclusive. Moreover,

$$(E \cap F_1) \cup (E \cap F_2) = E.$$

(This subdivision of E into mutually exclusive events is called a partition of E. See the discussion of the term "partition" in Section 12.2.) It follows that

$$P(E) = P(E \cap F_1) + P(E \cap F_2).$$

We now use (5) to compute $P(E \cap F_1)$, $P(E \cap F_2)$. Since the urn is selected at random, we have $P(F_1) = \frac{1}{2}$, and, assuming a random draw, we have $P(E \mid F_1) = \frac{3}{8}$; hence, $P(E \cap F_1) = P(F_1)P(E \mid F_1) = (\frac{1}{2})(\frac{3}{8}) = \frac{3}{16}$ by (5). Similarly, $P(E \cap F_2) = P(F_2)P(E \mid F_2) = (\frac{1}{2})(\frac{1}{6}) = \frac{1}{12}$, and therefore $P(E) = \frac{3}{16} + \frac{1}{12} = \frac{13}{48}$.

It is possible to extend (5) to an intersection of three or more events. For the case of three events, we have

$$\begin{aligned} P(E \cap F \cap G) &= P(E \cap (F \cap G)) \\ &= P(F \cap G)P(E \mid (F \cap G)), \quad \text{by (5).} \end{aligned}$$

But $$P(F \cap G) = P(G)P(F \mid G), \quad \text{again by (5),}$$
and a substitution then gives us

$$P(E \cap F \cap G) = P(G)P(F \mid G)P(E \mid (F \cap G)). \tag{6}$$

Example 3. Let $G =$ "Sunday is sunny," $F =$ "Tom's club goes on a picnic," $E =$ "Tom gets sunburned." The weather forecast states that $P(G) = 0.8$, a poll of the club suggests that $P(F \mid G) = 0.7$, and Tom's parents say that $P(E \mid (F \cap G)) = 0.95$. Hence the probability that Sunday is sunny and the club goes on a picnic and Tom gets sunburned is $P(E \cap F \cap G) = (0.8)(0.7)(0.95) = 0.532$, by (6).

If $P(E) > 0$, we may interchange E and F in (5), thus obtaining $P(F \cap E) = P(E)P(F \mid E)$. But, since $E \cap F = F \cap E$, we have $P(E \cap F) = P(F \cap E)$, and therefore

$$P(E)P(F \mid E) = P(F \cap E) = P(E \cap F) = P(F)P(E \mid F) \tag{7}$$

provided $P(E) > 0, P(F) > 0$.

Exercises [A]

1. In an old survey involving 6800 men in the German state of Baden, the following data were recorded:

		Color of Hair				
		Blond	Brown	Black	Red	Total
Color of Eyes	Blue	1768	807	189	47	2811
	Grey/green	946	1387	746	53	3132
	Brown	115	438	288	16	857
	Total	2829	2632	1223	116	6800

(a) If a man were selected at random from this set, (i) what would be the probability of his having brown hair and brown eyes? (ii) what would be the probability of his having brown hair, given that he had brown eyes? (iii) what would be the probability of his having brown eyes, given that he had brown hair?

(b) If a man with brown hair and a man with red hair were each chosen at random, which would be more likely to have blue eyes?

2. A red die and a green die are rolled (Exercise 4, page 502).

E_1 = "Red die shows 2." E_4 = "Sum of numbers is 5."
E_2 = "At least one die shows 2." E_5 = "Red die shows even number."
E_3 = "Sum of numbers is 7." E_6 = "Sum of numbers is even."

(a) Compute $P(E_1 \mid E_3)$, $P(E_3 \mid E_1)$. Are E_1 and E_3 independent?

(b) Compute $P(E_2 \mid E_3)$, $P(E_3 \mid E_2)$. Are E_2 and E_3 independent?

(c) Compute $P(E_1 \mid E_4)$, $P(E_4 \mid E_1)$. Are E_1 and E_4 independent?

(d) Compute $P(E_2 \mid E_5)$, $P(E_5 \mid E_2)$.

(e) Compute $P(E_5 \mid E_6)$, $P(E_6 \mid E_5)$.

(f) Compute $P(E_1 \mid E_6)$, $P(E_2 \mid E_6)$.

3. Assume that each time a child is born, $P(\text{Boy}) = P(\text{Girl}) = \frac{1}{2}$, and that the sexes of several births to the same parents are independent events. Consider all families with two children, and compute the probability that both children are boys, given that (a) one of them is a boy, (b) the older child is a boy.

4. (a) An honest coin is tossed three times. Find (i) the probability that the third toss is heads, (ii) the conditional probability that the third toss is heads, given that the first two tosses were heads.

(b) The same coin is tossed N times, $N > 1$. Find (i) the probability that the Nth toss is heads, (ii) the conditional probability that the Nth toss is heads, given that the first $N - 1$ tosses were heads.

5. What can be said about $P(A \mid B)$ if (a) A and B are mutually exclusive? (b) $B \subseteq A$? (c) $A \subseteq B$?

6. Prove: If $P(F) > 0$, then (a) $P(\emptyset \mid F) = 0$, (b) $0 \leq P(E \mid F) \leq 1$.

7. Cards are dealt, one at a time, from a well-shuffled pack. Find the probability that (a) the 5th card is the first spade, (b) the 10th card is the first ace.

8. In Example 2, page 530:

   (a) Compute the probability of drawing (i) a red ball, (ii) a blue ball.

   (b) Compute the conditional probability that a ball came from urn A, given that it is white.

9. Abel and Baker call the toss on a fair coin, Abel winning the toss if the coin shows heads, Baker if it shows tails. They toss four times.

   (a) What is the probability that Abel wins exactly three times?

   (b) What is the conditional probability that Abel wins exactly three times, given that he wins at least one toss?

   (c) What is the conditional probability that Abel wins exactly three times, given that he wins the first toss?

10. An urn contains four white balls and two red ones. A sample of four balls is drawn, one at a time, without replacement. Compute the probability of each of the following events.

   (a) The third ball drawn is white.

   (b) The third ball drawn is white, given that the sample contains a total of three white balls.

11. Do Exercise 10 for the case of sampling with replacement.

12. Prove. If $P(E \mid F) = P(E)$ and $P(E) > 0$, then $P(F \mid E) = P(F)$.

Exercises [B]

1. Prove.

   (a) $P(A \mid E) + P(\overline{A} \mid E) = 1 \quad (P(E) > 0)$.

   (b) $P(A \mid E) + P(A \mid \overline{E}) \geq P(A) \quad (P(E) > 0, P(\overline{E}) > 0)$.

   (c) $P((A \cup B) \mid E) = P(A \mid E) + P(B \mid E) - P((A \cap B) \mid E) \quad (P(E) > 0)$.

2. An urn contains b black balls and w white balls. Two balls are drawn, one after the other, without replacement. Find the probabilities of the following events.
   (a) The first ball is white.
   (b) The second ball is white, given that the first is white.
   (c) The second ball is white, given that the first is black.
   (d) The second ball is white.
3. A pack of cards is formed by using the four kings and the four queens from a regular deck. Two cards are drawn from this pack. Compute the probability that both of these cards are black, given the following events.
   (a) One of them is black. (b) One of them is a spade.
   (c) One of them is the king of spades.
4. One of two coins is honest. The other has been loaded so that the probability of obtaining a head on any one toss is $\frac{2}{3}$. One of these coins is selected at random, and tossed. What is the probability of a head?
5. Urn A contains 2 red balls and 3 white balls, urn B contains 4 red balls and 1 white ball. A ball is chosen at random from A and placed in B. The balls in B are then mixed, and one of them is chosen at random. What is the probability that it is white?
6. One coin is honest, another has two heads. One of these coins is selected at random and tossed twice. What is the probability that both tosses are (a) heads? (b) tails?

15.6 Bayes' Theorem

In certain applications, our formula (*4*) for conditional probability assumes a special form and acquires a special name, *Bayes' Theorem.* We start with two examples.

Example 1. Two coins look alike, but while one of them is an honest coin, with $P(H) = P(T) = \frac{1}{2}$, the other has been cleverly weighted so that $P(H) = \frac{2}{3}$, $P(T) = \frac{1}{3}$. We choose a coin at random and toss it three times, obtaining heads each time. Given this event, what is the probability that we chose the weighted coin?

Solution: Let $E_1 =$ "Honest coin chosen," $E_2 =$ "Weighted coin chosen," $F =$ "Three heads obtained." Since the selection of the coin is random, at the start of the experiment we have $P(E_1) = P(E_2) = \frac{1}{2}$. In the special vocabulary of Bayes' Theorem, these are called the prior (or a priori) probabilities of the events E_1, E_2. We are asked to compute the probability $P(E_2 \mid F)$; this is called the posterior (or a posteriori) probability of E_2. Common sense suggests that the event F should tip the scales in favor of E_2, and we therefore expect to find $P(E_2 \mid F) > \frac{1}{2}$.

Since the posterior probability $P(E_2 \mid F)$ is a conditional probability, we use (4), which in this case takes the form

$$P(E_2 \mid F) = \frac{P(E_2 \cap F)}{P(F)}. \tag{8}$$

To evaluate $P(F)$, we must take into account the fact that three heads can be obtained with either coin, and we therefore partition F (see Example 2, page 530) into the events $E_1 \cap F =$ "Honest coin and three heads" and $E_2 \cap F =$ "Weighted coin and three heads."

Since $F = (E_1 \cap F) \cup (E_2 \cap F)$ and $(E_1 \cap F) \cap (E_2 \cap F) = \emptyset$, we have

$$P(F) = P(E_1 \cap F) + P(E_2 \cap F),$$

and (8) therefore becomes

$$P(E_2 \mid F) = \frac{P(E_2 \cap F)}{P(E_1 \cap F) + P(E_2 \cap F)}. \tag{9}$$

Finally, to evaluate the probabilities on the right-hand side of (9), we turn to (5):

$$P(E_1 \cap F) = P(E_1)P(F \mid E_1),$$
$$P(E_2 \cap F) = P(E_2)P(F \mid E_2).$$

Substituting these results in (9) gives us

$$P(E_2 \mid F) = \frac{P(E_2)P(F \mid E_2)}{P(E_1)P(F \mid E_1) + P(E_2)P(F \mid E_2)}. \tag{10}$$

When our formula (4) for conditional probability assumes a form like that of (10), it is called Bayes' Theorem. We now evaluate (10).

We already have $P(E_1) = P(E_2) = \frac{1}{2}$. Assuming that successive tosses of a coin are independent, we find that the probability of obtaining three heads with the honest coin is $P(F \mid E_1) = (\frac{1}{2})^3 = \frac{1}{8}$ and that the probability of obtaining three heads with the weighted coin is $P(F \mid E_2) = (\frac{2}{3})^3 = \frac{8}{27}$. Hence (10) becomes

$$P(E_2 \mid F) = \frac{(\frac{1}{2})(\frac{8}{27})}{(\frac{1}{2})(\frac{1}{8}) + (\frac{1}{2})(\frac{8}{27})} = \frac{\frac{4}{27}}{\frac{91}{432}} = \frac{64}{91}.$$

Thus we see that taking the event F into account increases the probability of E_2 from its prior value of $P(E_2) = \frac{1}{2} = 0.500$ to its posterior value of $P(E_2 \mid F) = \frac{64}{91} \approx 0.703$.

There is also a posterior probability of E_1:

$$P(E_1 \mid F) = \frac{P(E_1 \cap F)}{P(F)} = \frac{P(E_1)P(F \mid E_1)}{P(E_1)P(F \mid E_1) + P(E_2)P(F \mid E_2)}$$
$$= \frac{(\frac{1}{2})(\frac{1}{8})}{(\frac{1}{2})(\frac{1}{8}) + (\frac{1}{2})(\frac{8}{27})} = \frac{\frac{1}{16}}{\frac{91}{432}} = \frac{27}{91}.$$

Because $E_1 = \overline{E_2}$, it follows from Exercise 1(a), page 532, that we should have $P(E_1 \mid F) + P(E_2 \mid F) = 1$. Since $\frac{27}{91} + \frac{64}{91} = 1$, we see that our results are consistent.

Example 2. Three urns contain red and white balls as shown.

An urn is chosen at random, and a ball is withdrawn. Find the posterior probabilities of having chosen each of the three urns, given that the ball is red.

	R	W	Total
Urn I	1	4	5
Urn II	1	2	3
Urn III	5	1	6

Solution: Let $E_1 =$ "Urn I is chosen," $E_2 =$ "Urn II is chosen," $E_3 =$ "Urn III is chosen," $F =$ "Ball drawn is red." We require $P(E_1 \mid F)$, $P(E_2 \mid F)$, $P(E_3 \mid F)$. For the first of these, we have

$$P(E_1 \mid F) = \frac{P(E_1 \cap F)}{P(F)} \tag{11}$$

from *(4)*, with similar formulas for the other two. Since a red ball can be drawn from any one of the three urns, the computation of $P(F)$ must take into account the three possibilities $E_1 \cap F$, $E_2 \cap F$, $E_3 \cap F$. These events are mutually exclusive in pairs and their union is F, so that by Theorem 2

$$P(F) = P(E_1 \cap F) + P(E_2 \cap F) + P(E_3 \cap F).$$

We again turn to *(5)* to evaluate the three probabilities on the right-hand side of this equation:

$$\begin{aligned} P(E_1 \cap F) &= P(E_1)P(F \mid E_1) = (\tfrac{1}{3})(\tfrac{1}{5}) = \tfrac{1}{15}, \\ P(E_2 \cap F) &= P(E_2)P(F \mid E_2) = (\tfrac{1}{3})(\tfrac{1}{3}) = \tfrac{1}{9}, \\ P(E_3 \cap F) &= P(E_3)P(F \mid E_3) = (\tfrac{1}{3})(\tfrac{5}{6}) = \tfrac{5}{18}. \end{aligned}$$

Substituting these results in *(11)*, we find

$$P(E_1 \mid F) = \frac{\frac{1}{15}}{\frac{1}{15} + \frac{1}{9} + \frac{5}{18}} = \tfrac{6}{41}.$$

Similarly, we have

$$\begin{aligned} P(E_2 \mid F) &= \frac{P(E_2 \cap F)}{P(F)} = \frac{\frac{1}{9}}{\frac{1}{15} + \frac{1}{9} + \frac{5}{18}} = \tfrac{10}{41}, \\ P(E_3 \mid F) &= \frac{P(E_3 \cap F)}{P(F)} = \frac{\frac{5}{18}}{\frac{1}{15} + \frac{1}{9} + \frac{5}{18}} = \tfrac{25}{41}. \end{aligned}$$

Note that for $i = 1, 2, 3$:

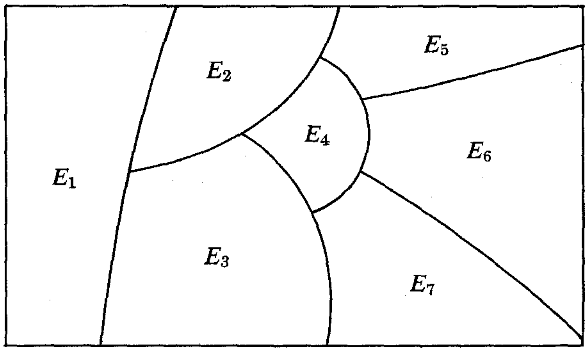
$$\begin{aligned} P(E_i \mid F) &= \frac{P(E_i)P(F \mid E_i)}{P(E_1)P(F \mid E_1) + P(E_2)P(F \mid E_2) + P(E_3)P(F \mid E_3)} \\ &= \frac{P(E_i)P(F \mid E_i)}{\sum_{j=1}^{3} P(E_j)P(F \mid E_j)}. \end{aligned}$$

This is again the Bayes' Theorem form of our formula *(4)* for conditional probability.

The general argument goes as follows: In a sample space S, we have n events $E_1, E_2, \cdots, E_n$, mutually exclusive in pairs, whose union is S:

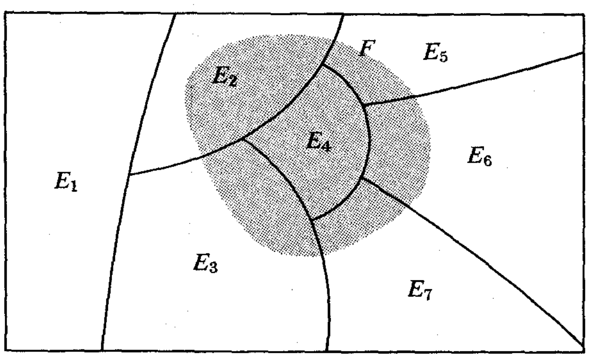
$$i \neq j \longrightarrow E_i \cap E_j = \emptyset,$$
$$E_1 \cup E_2 \cup \cdots \cup E_n = S.$$

As we stated in Example 2, page 530, a set of events $\{E_1, E_2, \cdots, E_n\}$ satisfying these conditions is called a *partition* of S. A partition of a set divides it



A Partition

into disjoint subsets just as the partitions in a house divide it into rooms. The figure above illustrates the case $n = 7$. If, now, F is any subset of S, then the set of events $\{E_1 \cap F, E_2 \cap F, \cdots, E_n \cap F\}$ is a partition of F: these events are disjoint, and their union is F. (See the figure below, in



A Partition of F

which the intersections $E_i \cap F$ are shaded.) Hence, by Theorem 2,

$$P(F) = \sum_{j=1}^{n} P(E_j \cap F).$$

But $P(E_j \cap F) = P(E_j)P(F \mid E_j)$ by *(5)*, so that

$$P(F) = \sum_{j=1}^{n} P(E_j)P(F \mid E_j). \qquad (12)$$

Now, for any integer i, $1 \leq i \leq n$, we have, by *(4)*,

$$P(E_i \mid F) = \frac{P(E_i \cap F)}{P(F)},$$

and substituting the results of *(5)* and *(12)* in this formula yields Bayes' Theorem:

$$P(E_i \mid F) = \frac{P(E_i)P(F \mid E_i)}{\sum_{j=1}^{n} P(E_j)P(F \mid E_j)}, \; 1 \leq i \leq n. \qquad (13)$$

The student should keep in mind that this rather frightening array of symbols is just a special way of writing the familiar formula *(4)* for conditional probability.

In using *(13)*, the events E_i are often called hypotheses or causes, and, as has been noted, the probability $P(E_i)$ is the prior or *a priori* probability, and the probability $P(E_i \mid F)$ the posterior or *a posteriori* probability, of the hypothesis E_i.

Example 3. Three automatic machines in a factory make bolts. Machine A, which is new, makes 10,000 bolts a day, and an average of 0.1% of them are defective. Machine B, which is older, makes 5,000 a day, with an average of 0.2% defective, and machine C, oldest of all, makes 5,000 a day, with an average of 0.3% defective. A day's output is mixed together, and a bolt is selected at random and found to be defective. What is the probability that it was made by C?

Solution: The sample space for the experiment of selecting a bolt at random contains 20,000 points, one corresponding to each bolt in the day's output. The hypotheses E_1 = "Bolt made by A," E_2 = "Bolt made by B," and E_3 = "Bolt made by C" are the elements of a partition of S since each sample point belongs to one and only one of these sets. Let F = "Bolt is defective." Then, by *(13)*,

$$\begin{aligned} P(E_3 \mid F) &= \frac{P(E_3)P(F \mid E_3)}{P(E_1)P(F \mid E_1) + P(E_2)P(F \mid E_2) + P(E_3)P(F \mid E_3)} \\ &= \frac{(0.25)(0.003)}{(0.5)(0.001) + (0.25)(0.002) + (0.25)(0.003)} \\ &= \tfrac{3}{7}. \end{aligned}$$

Exercises [A]

1. In Example 1, page 533, find the posterior probabilities of the two coins, given that all three tosses show tails.

2. In Example 2, page 535, find the posterior probabilities of the three urns, given that the ball drawn is white.

3. (Bertrand's Box.) A chest has three drawers. One contains two gold coins, one contains a gold coin and a silver coin, and one contains two silver coins. A drawer is selected at random, and a coin is removed from it. The coin is gold. What is the probability that the other coin in the same drawer is also gold?

4. Jane can pack 250 dozen eggs an hour into cartons, but she breaks about 1 in every 200 eggs. Martha only packs 200 dozen an hour, but she breaks only 1 in 500 eggs. Both girls are packing, and an egg is broken. What is the probability that Jane broke it?

5. **(a)** An honest die is tossed. If the 1 comes up, a ball is drawn from urn A. If any other number turns up, a ball is drawn from urn B. Urn A contains 4 white balls and 1 black ball. Urn B contains 2 white balls and 3 black balls. Find the posterior probabilities of the two urns if the ball drawn is white.

   (b) The ball drawn is returned to its urn, and a second random drawing is made from the same urn. It also is white. Find now the posterior probabilities of the two urns.

   (c) The second ball is replaced, and a third drawing is made from the same urn. If the third ball also is white, what now are the posterior probabilities of the two urns? Note how the accumulation of evidence affects the probabilities.

6. Students of probability are assigned in equal numbers to three instructors, Smith, Jones, and Brown. Smith and Jones give 5% A's, but Brown, an easy grader, gives 20% A's. Compute the probability that a particular student was assigned to Brown, given that **(a)** he got an A. **(b)** he did not get an A.

7. Each of three black boxes contains 1 green ball and 5 red balls. Each of two white boxes contains 3 green balls and 3 red balls. One of these boxes is chosen at random, and a ball is removed from it. The ball is red. What is the probability that the box chosen was black?

8. A card is lost from a full pack. Find the probability that the missing card is red, given each of the following events.

(a) A card drawn from the remaining pack is black.

(b) Two cards are drawn and both are black.

(c) Thirteen cards are drawn and all are black.

9. If 5% of all men and 0.25% of all women are colorblind, and a person chosen at random is found to be colorblind, what is the probability that the person chosen was a man?

10. An urn contains 3 black balls and 5 red balls. A ball is drawn at random and is then returned to the urn together with an additional ball of the same color. A second drawing is made. Compute the probability of each of the following events.

(a) The first ball drawn is black.

(b) The second ball drawn is black, given that the first ball is black.

(c) The second ball is black.

(d) The first ball is black, given that the second ball is black.

Exercises [B]

1. (See Exercise 6, page 533.) One of two coins is honest, the other has two heads. One of these coins is selected at random and tossed. It shows heads. What is the probability that another toss with the same coin will also show heads?

2. One card is drawn from a full deck. If the card is the ace of spades, a double-headed coin is tossed. If any other card is chosen, an ordinary coin is tossed. You do not know which coin was used, but you observe that it has been tossed n times and has shown a head every time.

(a) Express in terms of n the probability that the coin selected has two heads.

(b) For what values of n is this probability greater than $\frac{1}{2}$?

3. A number X is chosen at random from the set $N = \{1, 2, 3, 4, 5\}$. All elements of N greater than X are then discarded, and a second number Y is chosen from those that remain. (For example, if $X = 3$, then Y is chosen from $\{1, 2, 3\}$.)

(a) For each possible value of X, find the probability that $Y = 2$.

(b) Find the probability that $Y = 2$.

(c) Find the probability that $X = 3$, given that $Y = 2$.

4. An urn contains N balls, numbered consecutively from 1 to N. The value of N is not known, but it is known to be either 5 or 6 or 7, each with prior probability $\frac{1}{3}$. A sample of three balls is drawn from the urn, none numbered greater than 5. Find the posterior probabilities of the three possible values of N if the sample is drawn (a) with replacement (each ball is returned before the next one is drawn), (b) without replacement.

5. In a case of simple Mendelian inheritance, a gene may take a dominant (D) or a recessive (r) form. Each individual has two of these genes. If both are dominant (DD) or one is dominant and one recessive (Dr), then the individual shows dominant characteristics. If both are recessive (rr), then the individual shows recessive characteristics. When two individuals are crossed, the offspring receives one gene from each parent, and in the case of more than one offspring their genetic inheritances are independently determined. For example, if a Dr male is crossed with a Dr female, each offspring has one chance in four of being DD, two chances in four of being Dr, and one chance in four of being rr, and if a DD male is crossed with a Dr female, each offspring has two chances in four of being DD and two chances in four of being Dr. Now consider the following situation in which a male showing dominant characteristics (hence genetically either DD or Dr) is crossed with a female showing recessive characteristics (hence rr). If n offspring result, and all show dominant characteristics, what is the probability that the male parent is DD? Assume that the alternatives DD and Dr are equally likely, and evaluate for $n = 1, 2, 3, 4$.

Chapter Review

1. A die is made in the shape of a regular tetrahedron (triangular pyramid). The four faces are numbered 1, 2, 3, 4. When the die is rolled, the significant number is the one on the bottom. An experiment consists of rolling two such dice, one red and one green, and observing the ordered pair (r, g) = (number on red die, number on green die).
   (a) Determine a sample space for this experiment. Assign probabilities to the elementary events, assuming that the dice are not biased.
   (b) Determine the probability of each of the following events.
   $A = \{(r, g) : r + g \text{ is even}\}$
   $B = \{(r, g) : r + g = 4\}$
   $C = \{(r, g) : r = 3\}$
   $D = \{(r, g) : r > g\}$

(c) Compute $P(A \cap C)$, $P(A \cup C)$, $P(A \cup D)$, $P(B \mid D)$, $P(D \mid C)$, $P(A \mid (C \cup D))$.

2. Three cards are chosen at random, without replacement, from the A, K, Q, J, 10, 9 of spades. Compute the following probabilities.

(a) $P(A$ is chosen$)$

(b) $P(A$ is chosen $\mid K$ is chosen$)$

(c) $P(A$ is chosen or K is chosen$)$

(d) $P(A$ and K are chosen$)$

(e) $P(A$ is chosen $\mid K$ is not chosen$)$

(f) $P(A$ is chosen or K is chosen $\mid Q$ is not chosen or J is not chosen$)$

3. Five nickels and three dimes are distributed, one at a time, one coin to each of eight children. What is the probability that the sixth child gets the third dime?

4. An honest coin is tossed. If it shows heads, an ordinary six-sided die is tossed. If it shows tails, one of the tetrahedral dice of Exercise 1, above, is tossed. Let $E =$ "Number obtained is greater than 3," $T =$ "Coin shows tails." Compute (a) $P(E)$, (b) $P(T \mid E)$.

5. Each of two bags contains N nickels and N pennies. Prove that a random selection of one coin from each bag is more likely to yield two nickels than is mixing the contents of the two bags and then making a random selection of two coins.

6. A card player holds an ace and four other cards, none of them aces. He retains his ace, discards the other cards, and is dealt four new cards to go with the ace. What is the probability that he now has a pair of aces (but no more than two aces)?

Pages 449–451

1. (a) 8 (b) k **3.** (a) 90 (b) $k^2 + 3k + 2$
5. (a) 10 (b) 1 (c) 15 (d) 190 (e) 190

Pages 454–455

1. $1 + 5x + 10x^2 + 10x^3 + 5x^4 + x^5$
3. $1 - 12x + 60x^2 - 160x^3 + 240x^4 - 192x^5 + 64x^6$
5. $81x^4 + 108x^3y + 54x^2y^2 + 12xy^3 + y^4$
7. $1 - 7\sqrt{x} + 21x - 35x\sqrt{x} + 35x^2 - 21x^2\sqrt{x} + 7x^3 - x^3\sqrt{x}$
9. (a) $1 + 12x + 66x^2 + 220x^3 + 495x^4 + \cdots$ (b) 1.127 (approx.)
11. $x^8 + 16x^7y + 112x^6y^2 + 448x^5y^3 + \cdots$ **13.** $\frac{15}{128}$ **15.** $\frac{63}{256}$

Pages 457–458

1. (a) $1 + \frac{1}{2}x - \frac{1}{8}x^2 + \frac{1}{16}x^3 - \cdots$ (b) 1.0198 (approx.)
3. (a) $1 - 2x + 3x^2 - 4x^3 + \cdots$ (b) 1.0203 (approx.)
7. $1 + \frac{1}{4} + \frac{3}{32} + \frac{5}{128} + \cdots$

Pages 461–463

3. Convergent **5.** Convergent **7.** 2.718 (approx.) **9.** 1.0152 (approx.)
13. -0.223 (approx.) **15.** 0.09983 (approx.) **17.** $1 - \frac{x^2}{3!} + \frac{x^4}{5!} - \frac{x^6}{7!} + \cdots$
19. (b) -0.22307, 0.18227 (c) -0.32843, 0.36454, 0.69297

Chapter Review, Pages 463–466

1. (a) $-\frac{5}{2}$ (b) -35 **3.** (a) $\frac{9[1 - (\frac{1}{3})^n]}{2}$ (b) $\frac{9}{2}$ **5.** (a) $\frac{7}{9}$ (b) $\frac{17}{45}$
9. (a) 0 (b) 1 (c) 0 (d) none (e) 3 (f) none **11.** $|x| < 2$; $\frac{x}{2 - x}$
13. (a) yes (b) no **15.** (a) $\frac{1}{3}$ (b) no
17. (a) convergent, $\frac{4}{3}$ (b) divergent (c) convergent, $\frac{2}{3}$ (d) convergent, 1
19. $x^5 + 10x^4 + 40x^3 + 80x^2 + 80x + 32$
21. $2187x^7 - 10206x^6y + 20412x^5y^2 - 22680x^4y^3 + 15120x^3y^4 - 6048x^2y^5 + 1344xy^6 - 128y^7$ **23.** (a) $1 - \frac{1}{3}x + \frac{2}{9}x^2 - \frac{14}{81}x^3 + \cdots$ (b) 1.712 (approx.)
25. (a) convergent (b) divergent

Pages 472–473

1. 210 **3.** 24 **5.** 108 **7.** 12 **9.** (a) 60 (b) 125 **11.** (a) 125 (b) n^r
13. (a) 96 (b) 48 **15.** 1728 **17.** (a) 105 (b) 2 (c) 80 **19.** 6

Pages 476–477

1. 23 **3.** 271 **5.** 271 **7.** (a) 16 (b) 12 **9.** 265 **11.** 8190 **13.** 59
15. 104

Pages 483–484

1. (a) A (b) B **3.** $\emptyset$ **5.** 7
7. (a) 1, 2, 3, 4, 5, 6 (b) 1, 4 (c) 1, 2, 3, 5, 6, 8 (d) 5, 8

Pages 488–489

1. (a) 120 (b) 2450 (c) 100 **3.** 1 **5.** $\frac{20!}{11!}$ **7.** (a) 560 (b) 83160 **9.** 210
11. 5,652,770 **13.** 924 **15.** 2880 **17.** 56

Pages 489–490

1. (a) 21 (b) 21 (c) 21 **3.** (a) 120 (b) 120

Pages 494–496

1. 1326 **3.** 48620 **5.** (a) 20 (b) 10 (c) 6 (d) 4 (e) 4 (f) 16
7. (a) 462 (b) 200 (c) 30 **9.** 10 **11.** 4
13. (a) 220 (b) (i) 12 (ii) 4 (iii) 64 (iv) 144 (v) 64 **15.** 1296

Pages 496–497

1. 13,824 **3.** (a) 40 (b) 624 (c) 3744 (d) 5108 (e) 10,200 (f) 54,912
(g) 123,552 (h) 1,098,240 **5.** (a) 120 (b) 90 (c) 6

Chapter Review, Page 497

1. 1120 **3.** 1536

Pages 501–502

1. (a) (b) (f) **3.** (a) {(1, 2), (1, 3), (2, 1), (2, 3), (3, 1), (3, 2)}
(b) (i) {(2, 1), (3, 1), (3, 2)} (ii) {(3, 1), (3, 2)} (iii) {(1, 2), (3, 2)}
(iv) {(1, 3), (3, 1)} (v) {(2, 1), (3, 1), (3, 2)} (vi) {(3, 1), (3, 2)}
(vii) {(1, 2), (1, 3), (3, 1), (3, 2)} (viii) $\emptyset$ (ix) {(3, 2)} (x) {(3, 2)}

Pages 508–509

9. $P(E) = \frac{1}{4}$, $P(\overline{E}) = \frac{3}{4}$ **13.** (a) $\frac{1}{221}$ (b) $\frac{1}{17}$ (c) $\frac{1}{17}$ (d) $\frac{19}{34}$ (e) $\frac{33}{221}$
(f) $\frac{16}{17}$ (g) $\frac{29}{442}$ (h) $\frac{116}{221}$

Pages 516–518

3. (a) (i) $\frac{1}{8}$ (ii) $\frac{3}{8}$ (iii) $\frac{3}{8}$ (iv) $\frac{1}{8}$ (b) (i) yes (ii) yes (iii) no
7. (a) $\frac{4}{13}$ (b) $\frac{5}{52}$ (c) $\frac{11}{13}$ 9. (a) $\frac{4}{35}$ (b) $\frac{6}{7}$ 11. $\frac{13}{18}$
13. (a) $\{1, 2, 3, 4, 5, 6\}$; $P(1) = \frac{1}{21}$, $P(2) = \frac{2}{21}$, $P(3) = \frac{3}{21}$, $P(4) = \frac{4}{21}$, $P(5) = \frac{5}{21}$, $P(6) = \frac{6}{21}$ (b) $\frac{5}{7}$
15. (a) $\frac{7}{13}$ (b) $\frac{11}{13}$ (c) $\frac{7}{13}$

Pages 523–524

3. $P(E) = \frac{1}{3}$ and $P(F) = \frac{1}{8}$ or $P(E) = \frac{1}{8}$ and $P(F) = \frac{1}{3}$ 5. $\frac{1}{216}$
7. Not Independent 11. (a) $\frac{5}{54}$ (b) $\frac{671}{1296}$ (c) $\frac{5}{648}$
13. (a) 0.9857 (b) 0.0143 (c) 0.0000518

Pages 531–532

1. (a) (i) $\frac{219}{3400}$ (ii) $\frac{438}{857}$ (iii) $\frac{219}{1319}$ (b) Red-haired man 3. (a) $\frac{1}{3}$ (b) $\frac{1}{2}$
5. (a) $P(A \mid B) = 0$ (b) $P(A \mid B) = 1$ (c) $P(A \mid B) = \dfrac{P(A)}{P(B)}$
7. (a) 0.082 (b) 0.042 9. (a) $\frac{1}{4}$ (b) $\frac{4}{15}$ (c) $\frac{3}{8}$ 11. (a) $\frac{2}{3}$ (b) $\frac{2}{3}$

Pages 532–533

3. (a) $\frac{3}{11}$ (b) $\frac{5}{13}$ (c) $\frac{3}{7}$ 5. $\frac{7}{15}$

Pages 538–539

1. $P(E_1 \mid G) = \frac{27}{35}$, $P(E_2 \mid G) = \frac{8}{35}$ 3. $\frac{1}{2}$
5. (a) $P(A \mid W) = \frac{2}{7}$, $P(B \mid W) = \frac{5}{7}$ (b) $P(A \mid WW) = \frac{4}{9}$, $P(B \mid WW) = \frac{5}{9}$
(c) $P(A \mid WWW) = \frac{8}{13}$, $P(B \mid WWW) = \frac{5}{13}$
7. $\frac{5}{7}$ 9. $\frac{20}{21}$

Pages 539–540

1. $\frac{5}{6}$ 3. (a) $P(Y = 2 \mid X = 1) = 0$, $P(Y = 2 \mid X = 2) = \frac{1}{2}$, $P(Y = 2 \mid X = 3) = \frac{1}{3}$, $P(Y = 2 \mid X = 4) = \frac{1}{4}$, $P(Y = 2 \mid X = 5) = \frac{1}{5}$ (b) $\frac{77}{300}$ (c) $\frac{20}{77}$
5. $P(DD \mid D_n) = \dfrac{2^n}{2^n + 1}$; $P(DD \mid D_1) = \frac{2}{3}$, $P(DD \mid D_2) = \frac{4}{5}$, $P(DD \mid D_3) = \frac{8}{9}$, $P(DD \mid D_4) = \frac{16}{17}$

Chapter Review, Pages 540–541

1. (a) $\frac{1}{16}$ (b) $P(A) = \frac{1}{2}$; $P(B) = \frac{3}{16}$; $P(C) = \frac{1}{4}$; $P(D) = \frac{3}{8}$ (c) $P(A \cap C) = \frac{1}{8}$; $P(A \cup C) = \frac{5}{8}$; $P(A \cup D) = \frac{3}{4}$; $P(B \mid D) = \frac{1}{6}$; $P(D \mid C) = \frac{1}{2}$; $P(A \mid C \cup D) = \frac{3}{8}$ 3. $\frac{5}{28}$