probability measure if it satisfies the following three conditions.

(a) *P is finitely additive.*

(b) *P is nonnegative.*

(c) *$P(S) = 1$.*

In other words, for finite sample spaces probability is simply a measure which assigns the value 1 to the whole space.

It is important to realize that a complete description of a probability measure requires three things to be specified: the sample space S, the Boolean algebra $\mathscr{B}$ formed from certain subsets of S, and the set function P. The triple $(S, \mathscr{B}, P)$ is often called *a probability space.* In most of the elementary applications the Boolean algebra $\mathscr{B}$ is taken to be the collection of *all* subsets of S.

EXAMPLE. An illustration of applied probability theory is found in the experiment of tossing a coin once. For a sample space S we take the set of all conceivable outcomes of the experiment. In this case, each outcome is either "heads" or "tails," which we label by the symbols h and t. Thus, the sample space S is $\{h, t\}$, the set consisting of h and t. For the Boolean algebra we take the collection of all subsets of S; there are four, $\varnothing$, S, H, and T, where $H = \{h\}$ and $T = \{t\}$. Next, we assign probabilities to each of these subsets. For the subsets $\varnothing$ and S we have no choice in the assignment of values. Property (c) requires that $P(S) = 1$, and, since P is a nonnegative measure, $P(\varnothing) = 0$. However, there is some freedom in assigning probabilities to the other two subsets, H and T. Since H and T are disjoint sets whose union is S, the additive property requires that

$$P(H) + P(T) = P(S) = 1.$$

We are free to assign any nonnegative values whatever to $P(H)$ and $P(T)$ so long as their sum is 1. If we feel that the coin is unbiased so that there is no *a priori* reason to prefer heads or tails, it seems natural to assign the values

$$P(H) = P(T) = \tfrac{1}{2}.$$

If, however, the coin is "loaded," we may wish to assign different values to these two probabilities. For example, the values $P(H) = \frac{1}{3}$ and $P(T) = \frac{2}{3}$ are just as acceptable as $P(H) = P(T) = \frac{1}{2}$. In fact, for any real p in the interval $0 \le p \le 1$ we may define $P(H) = p$ and $P(T) = 1 - p$, and the resulting function P will satisfy all the conditions for a probability measure.

For a given coin, there is no mathematical way to determine what the probability p "really" is. If we choose $p = \frac{1}{2}$ we can deduce logical consequences on the assumption that the coin is fair or unbiased. The theory for unbiased coins can then be used to test the fairness of an actual coin by performing a large number of experiments with the coin and comparing the results with the predictions based on the theory. The testing of agreement between theory and empirical evidence belongs to that branch of applied probability known as *statistical inference,* and will not be discussed in this book.

The foregoing example is a typical application of the concepts of probability theory. Probability questions often arise in situations referred to as "experiments." We shall not attempt to define an experiment; instead, we shall merely mention some familiar examples: tossing one or more coins, rolling a pair of dice, dealing a bridge hand, drawing a ball from an urn, counting the number of female students at the California Institute of Technology, selecting a number from a telephone directory, recording the radiation count of a Geiger counter.

To discuss probability questions that arise in connection with such experiments, our first task is to construct a sample space S that can be used to represent all conceivable outcomes of the experiment, as we did for coin tossing. Each element of S should represent an outcome of the experiment and each outcome should correspond to one and only one element of S. Next, we choose a Boolean algebra $\mathscr{B}$ of subsets of S (usually *all* subsets of S) and then define a probability measure P on $\mathscr{B}$. The choice of the set S, the choice of $\mathscr{B}$, and the choice of P will depend on the information known about the details of the experiment and on the questions we wish to answer. The purpose of probability theory is not to discuss whether the probability space $(S, \mathscr{B}, P)$ has been properly chosen. This motivation belongs to the science or gambling game from which the experiment emanates, and only experience can suggest whether or not the choices were well made. *Probability theory is the study of logical consequences that can be derived once the probability space is given.* Making a good choice of the probability space is, strictly speaking, not probability theory — it is not even mathematics; instead, it is part of the art of applying probability theory to the real world. We shall elaborate further on these remarks as we deal with specific examples in the later sections.

If $S = \{a_1, a_2, \ldots, a_n\}$, and if $\mathscr{B}$ consists of all subsets of S, the probability function P is completely determined if we know its values on the one-element subsets, or *singletons:*

$$P(\{a_1\}), P(\{a_2\}), \ldots, P(\{a_n\}).$$

In fact, every subset A of S is a disjoint union of singletons, and $P(A)$ is determined by the additive property. For example, when

$$A = \{a_1\} \cup \{a_2\} \cup \cdots \cup \{a_k\},$$

the additive property requires that

$$P(A) = \sum_{i=1}^{k} P(\{a_i\}).$$

To simplify the notation and the terminology, we write $P(a_i)$ instead of $P(\{a_i\})$. This number is also called the *probability of the point* a_i. Therefore, the assignment of the point probabilities $P(x)$ for each element x in a finite set S amounts to a complete description of the probability function P.

13.6 Special terminology peculiar to probability theory

In discussions involving probability, one often sees phrases from everyday language such as "two events are equally likely," "an event is impossible," or "an event is certain to occur." Expressions of this sort have intuitive appeal and it is both pleasant and helpful to be able to employ such colorful language in mathematical discussions. Before we can do so, however, it is necessary to explain the meaning of this language in terms of the fundamental concepts of our theory.

Because of the way probability is used in practice, it is convenient to imagine that each probability space $(S, \mathscr{B}, P)$ is associated with a real or conceptual experiment. The universal set S can then be thought of as the collection of all conceivable outcomes of the experiment, as in the example of coin tossing discussed in the foregoing section. Each element of S is called an *outcome* or a *sample* and the subsets of S that occur in the Boolean algebra $\mathscr{B}$ are called *events*. The reasons for this terminology will become more apparent when we treat some examples.

Assume we have a probability space $(S, \mathscr{B}, P)$ associated with an experiment. Let A be an event, and suppose the experiment is performed and that its outcome is x. (In other words, let x be a point of S.) This outcome x may or may not belong to the set A. If it does, we say that *the event A has occurred*. Otherwise, we say that *the event A has not occurred,* in which case $x \in A'$, so the complementary event A' has occurred. An event A is called *impossible* if $A = \varnothing$, because in this case no outcome of the experiment can be an element of A. The event A is said to be *certain* if $A = S$, because then every outcome is automatically an element of A.

Each event A has a probability $P(A)$ assigned to it by the probability function P. [The actual value of $P(A)$ or the manner in which $P(A)$ is assigned does not concern us at present.] The number $P(A)$ is also called *the probability that an outcome of the experiment is one of the elements* of A. We also say that $P(A)$ is *the probability that the event A occurs* when the experiment is performed.

The impossible event $\varnothing$ must be assigned probability zero because P is a finitely additive measure. However, there may be events with probability zero that are not impossible. In other words, some of the nonempty subsets of S may be assigned probability zero. The certain event S must be assigned probability 1 by the very definition of probability, but there may be other subsets as well that are assigned probability 1. In Example 1 of Section 13.8 there are nonempty subsets with probability zero and proper subsets of S that have probability 1.

Two events A and B are said to be *equally likely* if $P(A) = P(B)$. The event A is called *more likely* than B if $P(A) > P(B)$, and *at least as likely* as B if $P(A) \geq P(B)$. Table 13.1 provides a glossary of further everyday language that is often used in probability discussions. The letters A and B represent events, and x represents an outcome of an experiment associated with the sample space S. Each entry in the left-hand column is a statement about the events A and B, and the corresponding entry in the right-hand column defines the statement in terms of set theory.

TABLE 13.1. Glossary of Probability Terms

Statement	*Meaning in set theory*
At least one of A or B occurs	$x \in A \cup B$
Both events A and B occur	$x \in A \cap B$
Neither A nor B occurs	$x \in A' \cap B'$
A occurs and B does not occur	$x \in A \cap B'$
Exactly one of A or B occurs	$x \in (A \cap B') \cup (A' \cap B)$
Not more than one of A or B occurs	$x \in (A \cap B)'$
If A occurs, so does B (A implies B)	$A \subseteq B$
A and B are mutually exclusive	$A \cap B = \varnothing$
Event A or event B	$A \cup B$
Event A and event B	$A \cap B$

13.7 Exercises

Let S be a given sample space and let A, B, and C denote arbitrary events (that is, subsets of S in the corresponding Boolean algebra $\mathscr{B}$). Each of the statements in Exercises 1 through 12 is described verbally in terms of A, B, C. Express these statements in terms of unions and intersections of A, B, C and their complements.

1. If A occurs, then B does not occur.
2. None of the events A, B, C occurs.
3. Only A occurs.
4. At least one of A, B, C occurs.
5. Exactly one of A, B, C occurs.
6. Not more than one occurs.
7. At least two of A, B, C occur.
8. Exactly two occur.
9. Not more than two occur.
10. A and C occur but not B.
11. All three events occur.
12. Not more than three occur.

13. Let A denote the event of throwing an odd total with two dice, and let B denote the event of throwing at least one 6. Give a verbal description of each of the following events:

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

14. Let A and B denote events. Show that

$$P(A \cap B) \leq P(A) \leq P(A \cup B) \leq P(A) + P(B).$$

15. Let A and B denote events and let $a = P(A)$, $b = P(B)$, $c = P(A \cap B)$. Compute, in terms of a, b, and c, the probabilities of the following events:

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

16. Given three events A, B, C. Prove that

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

13.8 Worked examples

We shall illustrate how some of the concepts of the foregoing sections may be used to answer specific questions involving probabilities.

EXAMPLE 1. What is the probability that at least one "head" will occur in two throws of a coin?

First Solution. The experiment in this case consists of tossing a coin twice; the set S of all possible outcomes may be denoted as follows:

$$S = \{hh, ht, th, tt\}.$$

If we feel that these outcomes are equally likely, we assign the point probabilities $P(x) = \frac{1}{4}$ for each x in S. The event "at least one head occurs" may be described by the subset

$$A = \{hh, ht, th\}.$$

The probability of this event is the sum of the point probabilities of its elements. Hence, $P(A) = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4}$.

Second Solution. Suppose we use the same sample space but assign the point probabilities as follows :†

$$P(hh) = 1, \qquad P(ht) = P(th) = P(tt) = 0.$$

Then the probability of the event "at least one head occurs" is

$$P(hh) + P(ht) + P(th) = 1 + 0 + 0 = 1 .$$

The fact that we arrived at a different answer from that in the first solution should not alarm the reader. We began with a different set of premises. Psychological considerations might lead us to believe that the assignment of probabilities in the first solution is the more natural one. Indeed, most people would agree that this is so if the coin is unbiased. However, if the coin happens to be loaded so that heads always turns up, the assignment of probabilities in the second solution is more natural.

The foregoing example shows that we cannot expect a unique answer to the question asked. To answer such a question properly we must specify the choice of sample space and the assignment of point probabilities. Once the sample space and the point probabilities are known only one probability for a given event can be logically deduced. Different choices of the sample space or point probabilities may lead to different "correct" answers to the same question.

Sometimes the assignment of probabilities to the individual outcomes of an experiment is dictated by the language used to describe the experiment. For example, when an object is chosen "at random" from a finite set of n elements, this is intended to mean that each outcome is equally likely and should be assigned point probability $1/n$. Similarly, when we toss a coin or roll a die, if we have no *apriori* reason to feel that the coin or die is loaded, we assume that all outcomes are equally likely. This agreement will be adopted in all the exercises of this chapter.

EXAMPLE 2. If one card is drawn at random from each of two decks, what is the probability that at least one is the ace of hearts?

Solution. The experiment consists in drawing two cards, a and b, one from each deck. Suppose we denote a typical outcome as an ordered pair (a, b). The number of possible outcomes, that is, the total number of distinct pairs (a, b) in the sample space S is 52^2. We assign the probability $1/52^2$ to each such pair. The event in which we are interested is the set A of pairs (a, b), where either a or b is the ace of hearts. There are 52 + 51 elements in A. Hence, under these assumptions we deduce that

$$P(A) = \frac{52 + 51}{52^2} = \frac{1}{26} - \frac{1}{52^2}.$$

† Note that for this assignment of probabilities there are nonempty subsets of S with probability zero and proper subsets with probability 1.

EXAMPLE 3. If two cards are drawn at random from one deck, what is the probability that one of them is the ace of hearts?

Solution. As in Example 2 we use ordered pairs (a, b) as elements of the sample space. In this case the sample space has $52 \cdot 51$ elements and the event A under consideration has $51 + 51$ elements. If we assign the point probabilities $1/(52 \cdot 51)$ to each outcome, we obtain

$$P(A) = \frac{2 \cdot 51}{52 \cdot 51} = \frac{1}{26}.$$

EXAMPLE 4. What is the probability of throwing 6 or less with three dice?

Solution. We denote each outcome of the experiment as a triple of integers (a, b, c) where a, b, and c may take any values from 1 to 6. Therefore the sample space consists of 6^3 elements and we assign the probability $1/6^3$ to each outcome. The event A in question is the set of all triples satisfying the inequality $3 \leq a + b + c \leq 6$. If A, denotes the set of (a, b, c) for which $a + b + c = n$, *we* have

$$A = A_3 \cup A_4 \cup A_5 \cup A_6.$$

Direct enumeration shows that the sets A,, with $n = 3, 4, 5,$ and 6 contain 1, 3, 6, and 10 elements, respectively. For example, the set A, is given by

$$A_6 = \{(1, 2, 3), (1, 3, 2), (1, 1, 4), (1, 4, 1), (2, 1, 3),$$
$$(2, 3, 1), (2, 2, 2), (3, 1, 2), (3, 2, 1), (4, 1, 1)\}.$$

Therefore A has 20 elements and

$$P(A) = \frac{20}{6^3} = \frac{5}{54}.$$

EXAMPLE 5. A die is thrown once. What is the probability that the number of points is either even or a multiple of 3 ?

Solution. We choose the sample space $S = \{1, 2, 3, 4, 5, 6\}$, consisting of six elements, to each of which we assign the probability $\frac{1}{6}$. The event "even" is the set $A = \{2, 4, 6\}$, the event "a multiple of 3" is B = (3, 6). We are interested in their union, which is the set $A \cup B$ = (2, 3, 4, 6). Since this set contains four elements we have $P(A \cup B) = 4/6$.

This example can be solved in another way, using the formula

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \tfrac{3}{6} + \tfrac{2}{6} - \tfrac{1}{6}.$$

13.9 Exercises

1. Let S be a finite sample space consisting of n elements. Suppose we assign equal probabilities to each of the points in S. Let A be a subset of S consisting of k elements. Prove that $P(A) = k/n$.

For each of Exercises 2 through 8, describe your choice of sample space and state how you are assigning the point probabilities. In the questions associated with card games, assume all cards have the same probability of being dealt.

2. Five counterfeit coins are mixed with nine authentic coins.
   (a) A coin is selected at random. Compute the probability that a counterfeit coin is selected.
   If two coins are selected, compute the probability that:
   (b) one is good and one is counterfeit.
   (c) both are counterfeit.
   (d) both are good.
3. Compute the probabilities of each of the events described in Exercise 13 of Section 13.7. Assign equal probabilities to each of the 36 elements of the sample space.
4. What is the probability of throwing at least one of 7, 11, or 12 with two dice?
5. A poker hand contains four hearts and one spade. The spade is discarded and one card is drawn from the remainder of the deck. Compute the probability of filling the flush-that is, of drawing a fifth heart.
6. In poker, a straight is a five-card sequence, not necessarily all of the same suit. If a poker hand contains four cards in sequence (but not A234 or JQKA) and one extra card not in the sequence, compute the probability of filling the straight. (The extra card is discarded and a new card is drawn from the remainder of the deck.)
7. A poker hand has four cards out of a five-card sequence with a gap in the middle (such as 5689), and one extra card not in the sequence. The extra card is discarded and a new one is drawn from the remainder of the deck. Compute the probability of filling the "inside straight."
8. An urn contains A white stones and B black stones. A second urn contains C white stones and D black stones. One stone is drawn at random from the first urn and transferred to the second urn. Then a stone is drawn at random from the second urn. Calculate the probability of each of the following events.
   (a) The first stone is white.
   (b) The first stone is black.
   (c) The second stone is white, given that the transferred stone was white.
   (d) The second stone is white, given that the transferred stone was black.
9. Two stones are drawn with replacement from an urn containing four red stones and two white stones. Calculate the probability of each of the following events.
   (a) Both stones are white.
   (b) Both stones are red.
   (c) Both stones are the same color.
   (d) At least one stone is red.
10. Let P_n denote the probability that exactly n of the events A and B will occur, where n takes the values 0, 1, 2. Express each of P_0, P_1, P_2 in terms of $P(A)$, $P(B)$, and $P(A \cap B)$.

Odds. Some gambling games are described in terms of "odds" rather than in terms of probabilities. For example, if we roll a fair die, the probability of the event "rolling a three" is $\frac{1}{6}$. Since there are six possible outcomes, one of which is favorable to the event "rolling a three" and five of which are unfavorable, this is often described by saying that the odds in favor of the event are 1 to 5, or the odds against it are 5 to 1. In this case the odds are related to the probability by the equation

$$\frac{1}{6} = \frac{1}{1 + 5}.$$

In general, if A is an event with probability $P(A)$ and if a and b are two real numbers such that

$$P(A) = \frac{a}{a + b}, \tag{13.6}$$

we say the ***odds in favor*** of A are a to b, or the ***odds against*** A are b to a. Since $1 - a/(a + b) = b/(a + b)$, the odds against A are the same as the odds in favor of the complementary event A'. The following exercises are devoted to further properties of odds and their relation to probabilities.

11. If $P(A) = 1$, show that (13.6) can be satisfied only when $b = 0$ and $a \neq 0$. If $P(A) \neq 1$, show that there are infinitely many choices of a and b satisfying (13.6) but that all have the same ratio a/b.
12. Compute the odds in favor of each of the events described in Exercise 2.
13. Given two events A and B. If the odds against A are 2 to 1 and those in favor of $A \cup B$ are 3 to 1, show that

$$\tfrac{5}{12} \leq P(B) \leq \tfrac{3}{4}.$$

Give an example in which $P(B) = \frac{5}{12}$ and one in which $P(B) = \frac{3}{4}$.

13.10 Some basic principles of combinatorial analysis

Many problems in probability theory and in other branches of mathematics can be reduced to problems on counting the number of elements in a finite set. Systematic methods for studying such problems form part of a mathematical discipline known as ***combinatorial analysis***. In this section we digress briefly to discuss some basic ideas in combinatorial analysis that are useful in analyzing some of the more complicated problems of probability theory.

If all the elements of a finite set are displayed before us, there is usually no difficulty in counting their total number. More often than not, however, a set is described in a way that makes it impossible or undesirable to display all its elements. For example, we might ask for the total number of distinct bridge hands that can be dealt. Each player is dealt 13 cards from a 52-card deck. The number of possible distinct hands is the same as the number of different subsets of 13 elements that can be formed from a set of 52 elements. Since this number exceeds 635 billion, a direct enumeration of all the possibilities is clearly not the best way to attack this problem; however, it can readily be solved by combinatorial analysis.

This problem is a special case of the more general problem of counting the number of distinct subsets of k elements that may be formed from a set of n elements,† where $n \geq k$. Let us denote this number by $f(n, k)$. It has long been known that

$$f(n, k) = \binom{n}{k}, \tag{13.7}$$

where, as usual $\binom{n}{k}$ denotes the binomial coefficient,

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}.$$

In the problem of bridge hands we have $f(52, 13) = \binom{52}{13} = 635{,}013{,}559{,}600$ different hands that a player can be dealt.

† When we say that a set has n elements, we mean that it has n ***distinct*** elements. Such a set is sometimes called an n-element set.

There are many methods known for proving (13.7). A straightforward approach is to form each subset of k elements by choosing the elements one at a time. There are n possibilities for the first choice, $n - 1$ possibilities for the second choice, and $n - (k - 1)$ possibilities for the kth choice. If we make all possible choices in this manner we obtain a total of

$$n(n - 1) \cdots (n - k + 1) = \frac{n!}{(n-k)!}$$

subsets of k elements. Of course, these subsets are not all distinct. For example, if $k = 3$ the six subsets

$$\{a, b, c\}, \{b, c, a\}, \{c, a, b\}, \{a, c, b\}, \{c, b, a\}, \{b, a, c\}$$

arc all equal. In general, this method of enumeration counts each k-element subset exactly $k!$ times.† Therefore we must divide the number $n!/(n - k)!$ by $k!$ to obtain $f(n, k)$. This gives us $f(n, k) = \binom{n}{k}$, as asserted.

This line of argument is more or less typical of the combinatorial analysis required in the later sections. Hence it seems worthwhile to digress briefly to discuss the fundamental principles on which this analysis is based.

We often wish to count the number of elements in the Cartesian product of n finite sets $A_1, \ldots, A_n$. The Cartesian product is denoted by the symbol $A_1 \times \cdots \times A_n$ and is defined by the equation

$$A_1 \times \cdots \times A_n = \{(a_1, \ldots, a_n) \mid a_1 \in A_1, \ldots, a_n \in A_n\}.$$

That is, the Cartesian product consists of the set of all ordered n-tuples $(a_1, \ldots, a_n)$ where the kth component of the n-tuple comes from the kth set A_k.

An example with $n = 2$ is shown in Figure 13.1. Here $A_1 = \{1, 2, 4, 5\}$ and $A_2 = \{1, 3\}$. There are 4 elements in A_1, and 2 elements in A_2, giving a total of 8 elements in the Cartesian product $A_1 \times A_2$. More generally, if A_1 consists of k_1 elements and A_2 consists of k_2 elements, then $A_1 \times A_2$ consists of $k_1 k_2$ elements. By induction on n, it follows that if A_r consists of k_r elements, then the Cartesian product $A_1 \times \cdots \times A_n$ consists of $k_1 \cdots k_n$ elements.

To express this result in the language of set functions, let $\mathscr{F}$ denote the class of all finite sets and let ν be the set function defined on $\mathscr{F}$ as follows: If $A \in \mathscr{F}$, $\nu(A)$ is the number of distinct elements in A. (For the empty set we define $\nu(\varnothing) = 0$.) Then it is easy to verify that ν is a finitely additive set function, so we may write

$$\nu\left(\bigcup_{i=1}^{n} S_i\right) = \sum_{i=1}^{n} \nu(S_i) \tag{13.8}$$

if $\{S_1, S_2, \ldots, S_n\}$ is a disjoint collection of finite sets (that is, if $S_i \cap S_j = \varnothing$ whenever $i \neq j$). The number of elements in a Cartesian product may be expressed in terms of ν as follows:

$$\nu(A_1 \times A_2 \times \cdots \times A_n) = \nu(A_1)\nu(A_2) \cdots \nu(A_n).$$

† The reason for this will become clear in Example 3 on p. 484, where we give a more detailed derivation of (13.7).

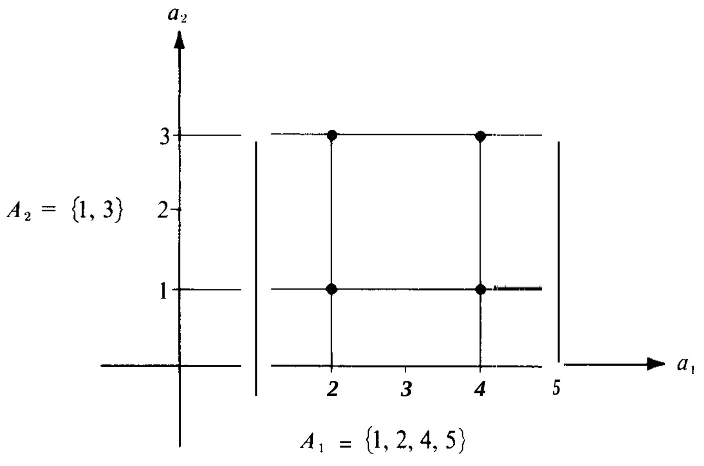

FIGURE 13.1 An example illustrating the Cartesian of two sets. The plotted points represent A, x A,.

A similar formula tells us how to count the number of elements in any set T of n-tuples if we know the number of possible choices for each of the successive components. For example, suppose there are k_1 possible choices for the first component x_1. Let k_2 be the number of possible choices for the second component x_2, once x_1 is known. Similarly, let k_r be the number of possible choices for the rth component x_r, once $x_1, x_2, \ldots, x_{r-1}$ are known. Then the number of n-tuples that can be formed with these choices is

$$\nu(T) = k_1 k_2 \ldots k, .$$

This formula is often referred to as the ***principle of sequential counting.*** It can be proved by induction on n. In many applications the set of choices for x_r may not be easy to describe since it may not be determined until after the choices of the earlier components have been made. (This was the case when we used the principle to count bridge hands.) Fortunately, to apply the principle of sequential counting we do not need to know the actual set of choices for x_r but only the ***number*** of possible choices for x_r.

The additive property in formula (13.8) and the principle of sequential counting provide the key to the solution of many counting problems. The following examples show how they can be applied.

EXAMPLE 1. ***Sampling with replacement.*** Given a set S consisting of n elements. If $k \geq 1$, how many ordered k-tuples can be formed if each component may be an arbitrary element of S?

Note: It may be helpful to think of S as an urn containing n balls labeled 1, 2, . . . , n. We select a ball and record its label as the first component of our k-tuple. Replacing the ball in the urn, we again select a ball and use its label as the second component, and so on, until we have made k selections. Since we replace each ball after it is drawn, the same label may appear in different components of our k-tuple.

Solution. Each k-tuple is an element of the Cartesian product

$$T = S_1 \times \cdots \times S_k,$$

where each $S_i = S$. Conversely, each element of T is one of the k-tuples in question. Hence the number of k-tuples formed in this way is

$$\nu(T) = \nu(S_1) \cdots \nu(S_k) = n^k.$$

EXAMPLE 2. *Sampling without replacement.* Given a set S consisting of n elements. If $k \le n$, how many ordered k-tuples can be formed if the components are chosen from S *without* replacement, that is to say, if no element of S may be used twice in any given k-tuple ?

Solution. Consider any k-tuple $(x_1, x_2, \ldots, x_k)$ formed from the elements of S without replacement. For the first component x_1 there are n choices (the n elements of S). When x_1 is chosen there remain $n - 1$ ways to choose x_2. With x_2 chosen, there remain $n - 2$ ways to choose x_3, and so on, there being $n - k + 1$ ways to choose x_k. Therefore, by the principle of sequential counting, the total number of k-tuples so formed is

$$n(n-1)(n-2)\cdots(n-k+1) = \frac{n!}{(n-k)!}.$$

In particular, when $k = n$ this result tells us that $n!$ distinct n-tuples may be formed from a given set S of n elements, with no two components of any n-tuple being equal.

EXAMPLE 3. *The number of k-element subsets of an n-element set.* If $k \le n$, how many distinct subsets of k elements can be formed from a set S consisting of n elements?

Solution. Let r denote the number of subsets in question and let us denote these subsets by

$$A_1, A_2, \ldots, A_r.$$

These sets are distinct but need not be disjoint. We shall compute r in terms of n and k by an indirect method. For this purpose, let B_i denote the collection of k-tuples that can be formed by choosing the components from the elements of A_i without replacement. The sets $B_1, B_2, \ldots, B_r$ *are* disjoint. Moreover, if we apply the result of Example 2 with $n = k$ we have

$$\nu(B_i) = k! \qquad \text{for each } i = 1, 2, \ldots, r.$$

Now let

$$T = B_1 \cup B_2 \cup \cdots \cup B_r.$$

Then T consists of all k-tuples that can be formed by choosing the components from S without replacement. From Example 2 we have

$$\nu(T) = n!/(n-k)!$$

and by additivity we also have

$$\nu(T) = \sum_{i=1}^{k!} \nu(B_i) = k!\, r.$$

Equating the two expressions for $\nu(T)$ we obtain

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

This proves formula (13.7) stated earlier in this section.

If we use the result of Example 3 to count the total number of subsets of a set S consisting of n elements, we obtain

$$\sum_{k=0}^{n} \binom{n}{k}.$$

Since this sum is also obtained when we expand $(1 + 1)^n$ by the binomial theorem, the number of subsets of S is 2^n.

13.11 Exercises

1. Let $A = \{1, 2, 3\}$. Display in roster notation the set of ordered pairs (a, b) obtained by choosing the first component from A and the second component from the *remaining* elements of A. Can this set of pairs be expressed as a Cartesian product?
2. A two-card hand can be dealt from a deck of 52 cards in $52 \cdot 51 = 2652$ ways. Determine the number of *distinct* hands, and explain your reasoning.
3. A senate committee consisting of six Democrats and four Republicans is to choose a chairman and a vice-chairman. In how many ways can this pair of officers be chosen if the chairman must be a Democrat ?
4. An experiment consists of tossing a coin twice and then rolling a die. Display each outcome of this experiment as an ordered triple (a, b, c), where each of a and b is either H (heads) or T (tails) and c is the number of points on the upturned face of the die. For example, $(H, H, 3)$ means that heads came up on both tosses and 3 appeared on the die. Express the set of all possible outcomes as a Cartesian product and determine the number of possible outcomes.
5. In how many ways can a bridge deck of 52 cards be dealt into four hands, each containing 13 cards? Explain your reasoning.
6. Two dice, one red and one white, are tossed. Represent the outcome as an ordered pair (a, b), where a denotes the number of points on the red die, b the number on the white die. How many ordered pairs (a, b) are possible ? How many are there for which the sum $a + b$ is:
   (a) even? (b) divisible by 3? (c) either even or divisible by 3?
7. A poker hand contains five cards dealt from a deck of 52. How many distinct poker hands can be dealt containing:
   (a) two pairs (for example, 2 kings, 2 aces, and a 3)?
   (b) a flush (five cards in a given suit)?
   (c) a straight flush (any five in sequence in a given suit, but not including ten, jack, queen, king, ace)?
   (d) a royal flush (ten, jack, queen, king, ace in a single suit)?

8. Refer to Exercise 7. Compute the probability for a poker hand to be:
   (a) a flush.
   (b) a straight flush.
   (c) a royal flush.
9. How many committees of 50 senators can be formed that contain:
   (a) exactly one senator from Alaska?
   (b) both senators from Alaska?
10. A committee of 50 senators is chosen at random. Compute the probability that both senators from Alaska are included.
11. A code group consists of four symbols in a row, each symbol being either a dot or a dash. How many distinct code groups can be formed?
12. How many k-letter words can be formed with an alphabet containing n letters?
13. Show that:

$$\text{(a)}\quad \binom{n}{0} + \binom{n}{2} + \binom{n}{4} + \cdots = \binom{n}{1} + \binom{n}{3} + \binom{n}{5} + \cdots = 2^{n-1}.$$

$$\text{(b)}\quad \binom{n}{0}^2 + \binom{n}{1}^2 + \quad + \binom{n}{n}^2 = \binom{2n}{n}.$$

14. Suppose a set of ordered pairs (a, b) is constructed by choosing the first component from a set of k elements, say from the set {a,, . . . , a,}, and the second component from a set of m elements, say from $\{b_1, \ldots, b_m\}$. There are a total of m pairs with first component a_1, namely, $(a_1, b_1), \ldots, (a_1, b_m)$. Similarly, there are m pairs $(a_i, b_1), \ldots, (a_i, b_m)$ with first component a_i. Therefore the total number of ordered pairs (a, b) is $m + m + \quad + m$ (k summands). This sum equals km, which proves the principle of sequential counting for sets of ordered pairs. Use induction to prove the principle for sets of ordered n-tuples.

13.12 Conditional probability

An unbiased die is thrown and the result is known to be an even number. What is the probability that this number is divisible by 3? What is the probability that a child is color blind, given that it is a boy? These questions can be put in the following form: Let A and B be events of a sample space S. What is the probability that A occurs, given that B has occurred? This is not necessarily the same as asking for the probability of the event $A \cap B$. In fact, when $A = B$ the question becomes: If A occurs, what is the probability that A occurs? The answer in this case should be 1 and this may or may not be the probability of $A \cap B$. To see how to treat such problems in general, we turn to the question pertaining to rolling a die.

When we ask probability questions about rolling an unbiased die, we ordinarily use for the sample space the set S = {1, 2, 3, 4, 5, 6) and assign point probability $\frac{1}{6}$ to each element of S. The event "divisible by 3" is the subset A = (3, 6) and the event "even" is the subset B = {2, 4, 6). We want the probability that an element is in A, given that it is in B. Since we are concerned only with outcomes in which the number is even, we disregard the outcomes 1, 3, 5 and use, instead of S, the set $B = \{2, 4, 6\}$ as our sample space. The event in which we are now interested is simply the singleton $\{6\}$, this being the only outcome of the new sample space that is divisible by 3. If all outcomes of B are considered equally likely, we must assign probability $\frac{1}{3}$ to each of them; hence, in particular, the probability of $\{6\}$ is also $\frac{1}{3}$.

Note that we solved the foregoing problem by employing a very elementary idea. We simply changed the sample space from S to B and provided a new assignment of probabilities. This example suggests a way to proceed in general.

Let $(S, \mathscr{B}, P)$ be a given probability space. Suppose A and B are events and consider the question: "What is the probability that A occurs, given that B has occurred?" As in the example just treated, we can change the sample space from S to B and provide a new assignment of probabilities. We are at liberty to do this in any manner consistent with the definition of probability measures. For B itself we have no choice except to assign the probability 1. Since we are interested in those elements of A which lie in the new sample space B, the problem before us is to compute the probability of the event $A \cap B$ according to the new assignment of probabilities. That is, if P' denotes the probability function associated with the new sample space B, then we must compute $P'(A \cap B)$.

We shall show now that if $P(B) \neq 0$ we can always define a probability function P' and a Boolean algebra $\mathscr{B}'$ of subsets of B such that $(B, \mathscr{B}', P')$ is a probability space. For the Boolean algebra $\mathscr{B}'$ we take the collection of all sets $T \cap B$ where T is in the original Boolean algebra $\mathscr{B}$. It is easy to verify that $\mathscr{B}'$, so defined, is indeed a Boolean algebra. One way to define a probability function P' on $\mathscr{B}'$ is simply to divide each of the old probabilities by $P(B)$. That is, if $C \in \mathscr{B}'$ we let

$$P'(C) = \frac{P(C)}{P(B)}.$$

(This is where the assumption $P(B) \neq 0$ comes in.) We are only changing the scale, with all probabilities magnified by the factor $1/P(B)$. It is easy to check that this definition of P' gives us a *bona jide* probability measure. It is obviously nonnegative and it assigns probability 1 to B. The additive property follows at once from the corresponding additive property for P.

Since each C in $\mathscr{B}'$ is of the form $A \cap B$, where A is an event in the original sample space S, we can rewrite the definition of P' as follows:

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

This discussion suggests that the quotient $P(A \cap B)/P(B)$ provides a reasonable measure of the probability that A occurs, given that B occurs. The following definition is made with this motivation in mind:

DEFINITION OF CONDITIONAL PROBABILITY. *Let $(S, \mathscr{B}, P)$ be a probability space and let B be an eoent such that $P(B) \neq 0$. The conditional probability that an event A will occur, given that B has occurred, is denoted by the symbol $P(A \mid B)$ (read: "the probability of A, given B") and is defined by the equation*

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

The conditional probability $P(A \mid B)$ is not defined if $P(B) = 0$.

The following examples illustrate the use of the concept of conditional probability.

EXAMPLE 1. Let us consider once more the problem raised earlier in this section: A die is thrown and the result is known to be an even number. What is the probability that this number is divisible by 3? As a problem in conditional probabilities, we may take for the sample space the set $S = \{1, 2, 3, 4, 5, 6\}$ and assign point probabilities $\frac{1}{6}$ to each element of S. The event "even" is the set $B = \{2, 4, 6\}$ and the event "divisible by 3" is the set $A = \{3, 6\}$. Therefore we have

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{1/6}{3/6} = \frac{1}{3}.$$

This agrees, of course, with the earlier solution in which we used B as the sample space and assigned probability $\frac{1}{3}$ to each element of B.

EXAMPLE 2. This is an example once used by the Caltech Biology Department to warn against the fallacy of superficial statistics. To "prove" statistically that the population of the U.S. contains more boys than girls, each student was asked to list the number of boys and girls in his family. Invariably, the total number of boys exceeded the total number of girls. The statistics in this case were biased because all undergraduates at Caltech were males. Therefore, the question considered here is not concerned with the probability that a child is a boy; rather, it is concerned with the conditional probability that a child is a boy, given that he comes from a family with at least one boy.

To compute the probabilities in an example of this type consider a sample of $4n$ families, each with two children. Assume that n families have 2 boys, $2n$ families one boy and one girl, and n families 2 girls. Let the sample space S be the set of all $8n$ children in these families and assign the point probability $P(x) = 1/(8n)$ to each x in S. Let A denote the event "the child is a boy" and B the event "the child comes from a family with at least one boy." The probability $P(A)$ is obviously $\frac{1}{2}$. Similarly, $P(B) = \frac{3}{4}$ since $3n$ of the $4n$ families have at least one boy. Therefore the probability that a child is a boy, given that he comes from a family with at least one boy, is the conditional probability

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)}{P(B)} = \frac{1/2}{3/4} = \frac{2}{3}.$$

13.13 Independence

An important idea related to conditional probability is the concept of *independence of events,* which may be defined as follows:

DEFINITION OF INDEPENDENCE. *Two events A and B are called independent (or stochastically independent) if and only if*

$$P(A \cap B) = P(A)P(B). \tag{13.9}$$

If A and B are independent, then $P(A \mid B) = P(A)$ if $P(B) \neq 0$. That is, the conditional probability of A, given B, is the same as the "absolute" probability of A. This relation exhibits the significance of independence. The knowledge that B has occurred does not influence the probability that A will occur.

EXAMPLE 1. One card is drawn from a 52-card deck. Each card has the same probability of being selected. Show that the two events "drawing an ace" and "drawing a heart" are independent.

Solution. We choose a sample space S consisting of 52 elements and assign the point probability $\frac{1}{52}$ to each element. The event A, "drawing an ace," has the probability $P(A) = \frac{4}{52} = \frac{1}{13}$. The event B, "drawing a heart," has the probability $P(B) = \frac{13}{52} = \frac{1}{4}$. The event $A \cap B$ means "drawing the ace of hearts," which has probability $\frac{1}{52}$. Since $P(A \cap B) = P(A)P(B)$, Equation (13.9) is satisfied and events A and B are independent.

EXAMPLE 2. Three true dice are rolled independently, so that each combination is equally probable. Let A be the event that the sum of the digits shown is six and let B be the event that all three digits are different. Determine whether or not these two events are independent.

Solution. For a sample space S we take the set of all triples (a, b, c) with a, b, c ranging over the values 1, 2, 3, 4, 5, 6. There are 6^3 elements in S, and since they are equally probable we assign the point probability $1/6^3$ to each element. The event A is the set of all triples (a, b, c) for which $a + b + c = 6$. Direct enumeration shows that there are 10 such triples, namely:

$$(1, 2, 3), (1, 3, 2), (1, 1, 4), (1, 4, 1),$$
$$(2, 1, 3), (2, 3, 1), (2, 2, 2),$$
$$(3, 1, 2), (3, 2, 1),$$
$$(4, 1, 1).$$

The event B consists of all triples (a, b, c) for which $a \neq b$, $b \neq c$, and $a \neq c$. There are $6 \cdot 5 \cdot 4 = 120$ elements in B. Exactly six of these elements are in set A, so that $A \cap B$ has six elements. Therefore

$$P(A \cap B) = \frac{6}{6^3}, \qquad P(A) = \frac{10}{6^3}, \qquad \text{and} \qquad P(B) = \frac{120}{6^3}.$$

In this case $P(A \cap B) \neq P(A)P(B)$; therefore events A and B are not independent.

Independence for more than two events is defined as follows. A finite collection $\mathscr{A}$ of n events is said to be independent if the events satisfy the multiplicative property

$$P\left(\bigcap_{k=1}^{m} A_k\right) = \prod_{k=1}^{m} P(A_k) \tag{13.10}$$

for every finite subcollection $\{A_1, A_2, \ldots, A_m\}$, where m may take the values $m = 2, 3, \ldots, n$, the sets A_i being in $\mathscr{A}$.

When $\mathscr{A}$ consists of exactly three events A, B, and C, the condition of independence in (13.10) requires that

$$P(A \cap B) = P(A)P(B), \; P(A \cap C) = P(A)P(C), \; P(B \cap C) = P(B)P(C), \tag{13.11}$$

and

$$P(A \cap B \cap C) = P(A)P(B)P(C). \tag{13.12}$$

It might be thought that the three equations in (13.11) suffice to imply (13.12) or, in other words, that independence of three events is a consequence of independence *in pairs*. This is not true, as one can see from the following example:

Four tickets labeled *a, b, c,* and *abc,* are placed in a box. A ticket is drawn at random, and the sample space is denoted by

$$S = \{a, b, c, abc\}.$$

Define the events A, B, and C as follows:

$$A = \{a, abc\}, \qquad B = \{b, abc\}, \qquad C = \{c, abc\}.$$

In other words, the event X means that the ticket drawn contains the letter x. It is easy to verify that each of the three equations in (13.11) is satisfied so that the events A, B, and C are independent in pairs. However, (13.12) is not satisfied and hence the *three* events are not independent. The calculations are simple and are left as an exercise for the reader.

13.14 Exercises

1. Let A and B be two events with $P(A) \neq 0$, $P(B) \neq 0$. Show that

$$P(A \cap B) = P(B)P(A \mid B) = P(A)P(B \mid A). \tag{13.13}$$

Sometimes it is easier to compute the probabilities $P(A)$ and $P(B \mid A)$ directly by enumeration of cases than it is to compute $P(A \cap B)$. When this is the case, Equation (13.13) gives a convenient way to calculate $P(A \cap B)$. The next exercise is an example.

2. An urn contains seven white and three black balls. A second urn contains five white and five black balls. A ball is selected at random from the first urn and placed in the second. Then a ball is selected at random from the second. Let A denote the event "black ball on first draw" and B the event "black ball on second draw."
(a) Compute the probabilities $P(A)$ and $P(B \mid A)$ directly by enumerating the possibilities. Use Equation (13.13) to compute $P(A \cap B)$.
(b) Compute $P(A \cap B)$ directly by enumerating all possible pairs of drawings.
3. (a) Let A_1, A_2, A_3 be three events such that $P(A_1 \cap A_2) \neq 0$. Show that

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1 \cap A_2).$$

(b) Use induction to generalize this result as follows: If $A_1, A_2, \ldots, A_n$ are n events $(n \geq 2)$ such that $P(A_1 \cap A_2 \cap \cdots \cap A_{n-1}) \neq 0$, then

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1 \cap A_2) \cdots P(A_n \mid A_1 \cap A_2 \cap \cdots \cap A_{n-1}).$$

4. A committee of 50 senators is chosen at random. Find the probability that both senators from Alaska are included, given that at least one is.

5. An urn contains five gold and seven blue chips. Two chips are selected at random (without replacement). If the first chip is gold, compute the probability that the second is also gold.
6. A deck of cards is dealt into four hands containing 13 cards each. If one hand has exactly eight spades, what is the probability that a particular one of the other hands has (a) at least one spade? (b) at least two spades? (c) a complete suit?
7. Show that $P(A \cup B \mid C) = P(A \mid C) + P(B \mid C) - P(A \cap B \mid C)$.
8. Let $A_1, A_2, \ldots, A_n$ be n disjoint events whose union is the entire sample space S. For every event E we have the equation

$$E = E \cap S = E \cap \bigcup_{i=1}^{n} A_i = \bigcup_{i=1}^{n} (E \cap A_i).$$

This equation states that E can occur only in conjunction with some A_i. Show that

(a) $P(E) = \sum_{i=1}^{n} P(E \cap A_i)$.

(b) $P(E) = \sum_{i=1}^{n} P(E \mid A_i)P(A_i)$.

This formula is useful when the conditional probabilities $P(E \mid A_i)$ are easier to compute directly than $P(E)$.
9. An unbiased coin is tossed repeatedly. It comes up heads on the first six tosses. What is the probability that it will come up heads on the seventh toss?
10. Given independent events A and B whose probabilities are neither 0 nor 1. Prove that A' and B' are independent. Is the same true if either of A or B has probability 0 or 1 ?
11. Given independent events A and B. Prove or disprove in each case that :
(a) A' and B are independent.
(b) $A \cup B$ and $A \cap B$ are independent.
(c) $P(A \cup B) = 1 - P(A')P(B')$.
12. If $A_1, A_2, \ldots, A_n$ are independent events, prove that

$$P\left(\bigcup_{i=1}^{n} A_i\right) + \prod_{i=1}^{n} P(A_i') = 1.$$

13. If the three events A, B, and C are independent, prove that $A \cup B$ and C are independent.

[*Hint:* Use the result of Exercise 7 to show that $P(A \cup B \mid C) = P(A \cup B)$.]

14. Let A and B be events, neither of which has probability 0. Prove or disprove the following statements :
(a) If A and B are disjoint, A and B are independent.
(b) If A and B are independent, A and B are disjoint.
15. A die is thrown twice, the sample space S consisting of the 36 possible pairs of outcomes (a, b) each assigned probability $\frac{1}{36}$. Let A, B, and C denote the following events:

$$A = \{(a, b) \mid a \text{ is odd}\}, \quad B = \{(a, b) \mid b \text{ is odd}\}, \quad C = \{(a, b) \mid a + b \text{ is odd}\}.$$

(a) Compute $P(A)$, $P(B)$, $P(C)$, $P(A \cap B)$, $P(A \cap C)$, $P(B \cap C)$, and $P(A \cap B \cap C)$.
(b) Show that A, B, and C are independent in pairs.
(c) Show that A, B, and C are not independent.

13.15 Compound experiments

We turn now to the problem of de Méré mentioned in the introduction -whether or not it is profitable to bet even money on the occurrence of at least one "double six" in 24 throws of a pair of dice. We treat the problem in a more general form: What is the probability of throwing a double six at least once in n throws of a pair of dice? Is this probability more than one-half or less than one-half when $n = 24$?

Consider first the experiment of tossing a pair of fair dice just once. The outcomes of this game can be described by ordered pairs (a, b) in which a and b range over the values 1, 2, 3, 4, 5, 6. The sample space S consists of 36 such pairs. Since the dice are fair we assign the probability $\frac{1}{36}$ to each pair in S.

Now suppose we roll the dice n times. The succession of the n experiments is one compound experiment which we wish to describe mathematically. To do this we need a new sample space and a corresponding probability measure. We consider the outcomes of the new game as ordered n-tuples $(x_1, \ldots, x_n)$, where each component x_i is one of the outcomes of the original sample space S. In other words, the sample space for the compound experiment is the n-fold Cartesian product $S \times \cdots \times S$, which we denote by S^n. The set S^n has 36^n elements, and we assign equal probabilities to each element:

$$P(x) = \frac{1}{36^n} \quad \text{if} \quad x \in S^n.$$

We are interested in the event "at least one double six in n throws." Denote this event by A. In this case it is easier to compute the probability of the complementary event A', which means "no double six in n throws." Each element of A' is an n-tuple whose components can be any element of S except (6, 6). Therefore there are 35 possible values for each component and hence $(35)^n$ n-tuples altogether in A'. Since each element of A' has probability $(\frac{1}{36})^n$, the sum of all the point probabilities in A' is $(\frac{35}{36})^n$. This gives us

$$P(A) = 1 - P(A') = 1 - (\tfrac{35}{36})^n .$$

To answer de Méré's question we must decide whether $P(A)$ is more than one-half or less than one-half when $n = 24$. The inequality $P(A) \geq \frac{1}{2}$ is equivalent to $1 - (\frac{35}{36})^n \geq \frac{1}{2}$, or $(\frac{35}{36})^n \leq \frac{1}{2}$. Taking logarithms we find

$$n \log 35 - n \log 36 \leq -\log 2, \qquad \text{or} \qquad n \geq \frac{\log 2}{\log 36 - \log 35} = 24.6+ .$$

Therefore $P(A) < \frac{1}{2}$ when $n = 24$ and $P(A) > \frac{1}{2}$ when $n \geq 25$. It is *not* advantageous to bet even money on the occurrence of at least one double six in 24 throws.

The foregoing problem suggests a general procedure for dealing with successive experiments. If an experiment is repeated two or more times, the result can be considered one compound experiment. More generally, a compound experiment may be the result of performing two or more distinct experiments successively. The individual experiments may be related to each other or they may be stochastically independent, in the sense that the probability of the outcome of any one of them is unrelated to the results of the others. For the sake of simplicity, we shall discuss how one can combine *two* independent experiments into one compound experiment. The generalization to more than two experiments will be evident.

To associate a *bona fide* probability space with a compound experiment we must explain how to define the new sample space S, the corresponding Boolean algebra $\mathscr{B}$ of subsets of S, and the probability measure P defined on $\mathscr{B}$. As in the above example, we use the concept of ***Cartesian product.***

Suppose we have two probability spaces, say (S,, $\mathscr{B}_1$, ***P,)*** and (S,, $\mathscr{B}_2$, P_2). These spaces may be thought of as associated with two experiments E_1 and E_2. By the compound experiment E we mean the one for which the sample space S is the Cartesian product $S_1 \times S_2$. An outcome of E is a pair (x, y) in S, with the first component x an outcome of E_1 and the second component y an outcome of E_2. If S_1 has n elements and if S_2 has m elements, then $S_1 \times S_2$ has nm elements.

For the new Boolean algebra $\mathscr{B}$ we take the collection of all subsets of S. Next we define the probability function P. Since S is finite we can define $P(x, y)$ for each point (x, y) in S and then use additivity to define P for subsets of S. The point probabilities $P(x, y)$ can be assigned in many ways. However, if the two experiments E_1 and E_2 are stochastically ***independent,*** we define P by the equation

$$P(x, y) = P_1(x)P_2(y) \qquad \text{for each } (x, y) \text{ in } S. \tag{13.14}$$

Motivation for this definition can be given as follows. Consider two special events A and B in the new space S,

$$A = \{(x_1, y_1), (x_1, y_2), \ldots, (x_1, y_m)\}$$

and

$$B = \{(x_1, y_1), (x_2, y_1), \ldots, (x_n, y_1)\}.$$

That is, A is the set of all pairs in $S_1 \times S_2$ whose first element is x_1, and B is the set of all pairs whose second element is y_1. The intersection of the two sets A and B is the singleton $\{(x_1, y_1)\}$. If we feel that the first outcome x_1 should have no influence on the second outcome y_1 it seems reasonable to require events A and B to be independent. This means we would like to define the new probability function P in such a way that we have

$$P(A \cap B) = P(A)P(B). \tag{13.15}$$

If we decide how to assign the probabilities $P(A)$ and $P(B)$, Equation (13.15) will tell us how to assign the probability $P(A \cap B)$, that is, the probability $P(x_1, y_1)$. Event A occurs if and only if the outcome of the first experiment is x_1. Since $P_1(x_1)$ is its probability, it seems natural to assign the value $P_1(x_1)$ to $P(A)$ as well. Similarly, we assign the value $P_2(y_1)$ to $P(B)$. Equation (13.15) then gives us

$$P(x_1, y_1) = P_1(x_1)P_2(y_1).$$

All this, of course, is merely motivation for the assignment of probabilities in (13.14). The only way to decide whether or not (13.14) is a permissible assignment of point probabilities is to check the fundamental properties of probability measures. Each number $P(x, y)$ is nonnegative, and the sum of all the point probabilities is equal to 1, since we have

$$\sum_{(x,y)\in S} P(x, y) = \sum_{x\in S_1} P_1(x) \cdot \sum_{y\in S_2} P_2(y) = 1 \cdot 1 = 1.$$