1

Basic Concepts

1.1 INTRODUCTION

The origin of probability theory lies in physical observations associated with games of chance. It was found that if an "unbiased" coin is tossed independently n times, where n is very large, the relative frequency of heads, that is, the ratio of the number of heads to the total number of tosses, is very likely to be very close to 1/2. Similarly, if a card is drawn from a perfectly shuffled deck and then is replaced, the deck is reshuffled, and the process is repeated over and over again, there is (in some sense) convergence of the relative frequency of spades to 1/4.

In the card experiment there are 52 possible outcomes when a single card is drawn. There is no reason to favor one outcome over another (the principle of "insufficient reason" or of "least astonishment"), and so the early workers in probability took as the probability of obtaining a spade the number of favorable outcomes divided by the total number of outcomes, that is, 13/52 or 1/4.

This so-called "classical definition" of probability (the probability of an event is the number of outcomes favorable to the event, divided by the total number of outcomes, where all outcomes are equally likely) is first of all restrictive (it considers only experiments with a finite number of outcomes) and, more seriously, circular (no matter how you look at it, "equally likely" essentially means "equally probable," and thus we are using the concept of probability to define probability itself). Thus we cannot use this idea as the basis of a mathematical theory of probability; however, the early probabilists were not prevented from deriving many valid and useful results.

Similarly, an attempt at a frequency definition of probability will cause trouble. If S_n is the number of occurrences of an event in n independent performances of an experiment, we expect physically that the relative frequency S_n/n should coverge to a limit; however, we cannot assert that the limit exists in a mathematical sense. In the case of the tossing of an unbiased coin, we expect that $S_n/n \to 1/2$, but a conceivable outcome of the process is that the coin will keep coming up heads forever. In other words it is possible that $S_n/n \to 1$, or that $S_n/n \to$ any number between 0 and 1, or that S_n/n has no limit at all.

In this chapter we introduce the concepts that are to be used in the construction of a mathematical theory of probability. The first ingredient we need is a set Ω, called the *sample space*, representing the collection of possible outcomes of a random experiment. For example, if a coin is tossed once we may take $\Omega = \{H, T\}$, where H corresponds to a head and T to a tail. If the coin is tossed twice, this is a different experiment and we need a different Ω, say $\{HH, HT, TH, TT\}$; in this case one performance of the experiment corresponds to two tosses of the coin.

If a single die is tossed, we may take Ω to consist of six points, say $\Omega = \{1, 2, \ldots, 6\}$. However, another possible sample space consists of two points, corresponding to the outcomes "N is even" and "N is odd," where N is the result of the toss. Thus different sample spaces can be associated with the same experiment. The nature of the particular problem under consideration will dictate which sample space is to be used. If we are interested, for example, in whether or not $N \geq 3$ in a given performance of the experiment, the second sample space, corresponding to "N even" and "N odd," will not be useful to us.

In general, the only physical requirement on Ω is that *a given performance of the experiment must produce a result corresponding to exactly one of the points of* Ω. We have as yet no mathematical requirements on Ω; it is simply a set of points.

Next we come to the notion of *event*. An "event" associated with a random experiment corresponds to a question about the experiment that has a yes or no answer, and this in turn is associated with a subset of the sample space. For example, if a coin is tossed twice and $\Omega = \{HH, HT, TH, TT\}$, "the number of heads is ≤ 1" will be a condition that either occurs or does not occur in a given performance of the experiment. That is, after the experiment is performed, the question "Is the number of heads ≤ 1?" can be answered yes or no. The subset of Ω corresponding to a "yes" answer is $A = \{HT, TH, TT\}$; that is, if the outcome of the experiment is HT, TH, or TT, the answer

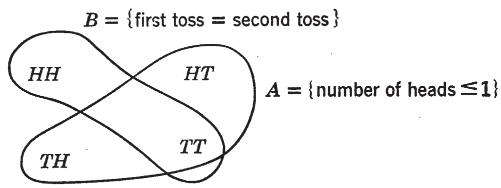


FIGURE 1.1.1 Coin-Tossing Experiment.

to the question "Is the number of heads ≤ 1?" will be "yes," and if the outcome is HH, the answer will be "no." Similarly, the subset of Ω associated with the "event" that the result of the first toss is the same as the result of the second toss is $B = \{HH, TT\}$.

Thus an *event* is defined as a subset of the sample space, that is, a collection of points of the sample space. (We shall qualify this in the next section.)

Events will be denoted by capital letters at the beginning of the English alphabet, such as A, B, C, and so on. An event may be characterized by listing all of its points, or equivalently by describing the conditions under which the event will occur. For example, in the coin-tossing experiment just considered, we write

$$A = \{\text{the number of heads is less than or equal to } 1\}$$

This expression is to be read as "A is the set consisting of those outcomes which satisfy the condition that the number of heads is less than or equal to 1," or, more simply, "A is the event that the number of heads is less than or equal to 1." The event A consists of the points HT, TH, and TT; therefore we write $A = \{HT, TH, TT\}$, which is to be read "A is the event consisting of the points HT, TH, and TT." As another example, if B is the event that the result of the first toss is the same as the result of the second toss, we may describe B by writing $B = \{\text{first toss} = \text{second toss}\}$ or, equivalently, $B = \{HH, TT\}$ (see Figure 1.1.1).

Each point belonging to an event A is said to be *favorable* to A. The event A will occur in a given performance of the experiment if and only if the outcome of the experiment corresponds to one of the points of A. The entire sample space Ω is said to be the *sure* (or *certain*) event; if *must* occur on any given performance of the experiment. On the other hand, the event consisting of none of the points of the sample space, that is, the empty set $\varnothing$, is called the *impossible event*; it can *never* occur in a given performance of the experiment.

1.2 ALGEBRA OF EVENTS (BOOLEAN ALGEBRA)

Before talking about the assignment of probabilities to events, we introduce some operations by which new events are formed from old ones. These operations correspond to the construction of compound sentences by use of the connectives "or," "and," and "not." Let A and B be events in the same sample space. Define the *union* of A and B (denoted by $A \cup B$) as the set consisting of those points belonging to *either A or B or both*. (Unless otherwise specified, the word "or" will have, for us, the inclusive connotation. In other words, the statement "p or q" will always mean "p or q or both.") Define the *intersection* of A and B, written $A \cap B$, as the set of points that belong to *both* A and B. Define the *complement* of A, written A^c, as the set of points which do *not* belong to A.

▶ **Example 1.** Consider the experiment involving the toss of a single die, with $N =$ the result; take a sample space with six points corresponding to $N = 1, 2, 3, 4, 5, 6$. For convenience, label the points of the sample space by the integers 1 through 6.

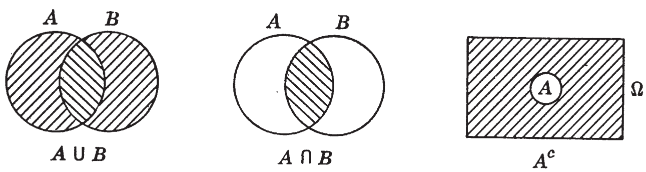


FIGURE 1.2.1 Venn Diagrams.

$$\text{Let } A = \{N \text{ is even}\} \qquad \text{and} \qquad B = \{N \geq 3\}$$

Then

$$A \cup B = \{N \text{ is even or } N \geq 3\} = \{2, 3, 4, 5, 6\}$$
$$A \cap B = \{N \text{ is even and } N \geq 3\} = \{4, 6\}$$
$$A^c = \{N \text{ is not even}\} = \{1, 3, 5\}$$
$$B^c = \{N \text{ is not } \geq 3\} = \{N < 3\} = \{1, 2\} \blacktriangleleft$$

Schematic representations (called *Venn diagrams*) of unions, intersections, and complements are shown in Figure 1.2.1.

Define the *union of n events* $A_1, A_2, \ldots, A_n$ (notation: $A_1 \cup \cdots \cup A_n$, or $\bigcup_{i=1}^n A_i$) as the set consisting of those points which belong to *at least one* of the events $A_1, A_2, \ldots, A_n$. Similarly define the union of an infinite sequence of events $A_1, A_2, \ldots$ as the set of points belonging to at least one of the events $A_1, A_2, \ldots$ (notation: $A_1 \cup A_2 \cup \cdots$, or $\bigcup_{i=1}^\infty A_i$).

Define the *intersection of n events* $A_1, \ldots, A_n$ as the set of points belonging to *all* of the events $A_1, \ldots, A_n$ (notation: $A_1 \cap A_2 \cap \cdots \cap A_n$, or $\bigcap_{i=1}^n A_i$). Similarly define the intersection of an infinite sequence of events as the set of points belonging to all the events in the sequence (notation: $A_1 \cap A_2 \cap \cdots$, or $\bigcap_{i=1}^{\infty} A_i$). In the above example, with $A = \{N \text{ is even}\} = \{2, 4, 6\}$, $B = \{N \geq 3\} = \{3, 4, 5, 6\}$, $C = \{N = 1 \text{ or } N = 5\} = \{1, 5\}$, we have

$$A \cup B \cup C = \Omega, \qquad A \cap B \cap C = \varnothing$$
$$A \cup B^c \cup C = \{2, 4, 6\} \cup \{1, 2\} \cup \{1, 5\} = \{1, 2, 4, 5, 6\}$$
$$(A \cup C) \cap [(A \cap B)^c] = \{1, 2, 4, 5, 6\} \cap \{4, 6\}^c = \{1, 2, 5\}$$

Two events in a sample space are said to be *mutually exclusive* or *disjoint* if A and B have no points in common, that is, if it is impossible that both A and B occur during the *same* performance of the experiment. In symbols, A and B are mutually exclusive if $A \cap B = \varnothing$. In general the events A_1, $A_2, \ldots, A_n$ are said to be mutually exclusive if no two of the events have a point in common; that is, no more than one of the events can occur during

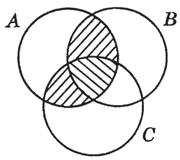


FIGURE 1.2.2 $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

the same performance of the experiment. Symbolically, this condition may be written

$$A_i \cap A_j = \varnothing \qquad \text{for } i \neq j$$

Similarly, infinitely many events $A_1, A_2, \ldots$ are said to be mutually exclusive if $A_i \cap A_j = \varnothing$ for $i \neq j$.

In some ways the algebra of events is similar to the algebra of real numbers, with union corresponding to addition and intersection to multiplication. For example, the commutative and associative properties hold.

$$A \cup B = B \cup A, \qquad A \cup (B \cup C) = (A \cup B) \cup C$$
$$A \cap B = B \cap A, \qquad A \cap (B \cap C) = (A \cap B) \cap C \qquad (1.2.1)$$

Furthermore, we can prove that for events A, B, and C in the same sample space we have

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \qquad (1.2.2)$$

There are several ways to establish this; for example, we may verify that the sets of both the left and right sides of the equality above are represented by the area in the Venn diagram of Figure 1.2.2.

Another approach is to use the definitions of union and intersection to show that the sets in question have precisely the same members; that is, we show that any point which belongs to the set on the left necessarily belongs to the set on the right, and conversely. To do this, we proceed as follows.

$$\begin{aligned} x \in A \cap (B \cup C) &\Rightarrow x \in A \quad \text{and} \quad x \in B \cup C \\ &\Rightarrow x \in A \quad \text{and} \quad (x \in B \text{ or } x \in C) \end{aligned}$$

(The symbol $\Rightarrow$ means "implies," and $\Leftrightarrow$ means "implies and is implied by.")

CASE 1. $x \in B$. Then $x \in A$ and $x \in B$, so $x \in A \cap B$, so $x \in (A \cap B) \cup (A \cap C)$.

CASE 2. $x \in C$. Then $x \in A$ and $x \in C$, so $x \in A \cap C$, so $x \in (A \cap B) \cup (A \cap C)$.

Thus $x \in A \cap (B \cup C) \Rightarrow x \in (A \cap B) \cup (A \cap C)$; that is, $A \cap (B \cup C) \subset (A \cap B) \cup (A \cap C)$. (The symbol $\subset$ is read "is a subset of"; we say that $A_1 \subset A_2$ provided that $x \in A_1 \Rightarrow x \in A_2$; see Figure 1.2.3. Notice that, according to this definition, a set A is a subset of itself: $A \subset A$.)

Conversely: Let $x \in (A \cap B) \cup (A \cap C)$. Then $x \in A \cap B$ or $x \in A \cap C$.

CASE 1. $x \in A \cap B$. Then $x \in B$, so $x \in B \cup C$, so $x \in A \cap (B \cup C)$.

CASE 2. $x \in A \cap C$. Then $x \in C$, so $x \in B \cup C$, so $x \in A \cap (B \cup C)$.

Thus $(A \cap B) \cup (A \cap C) \subset A \cap (B \cup C)$; hence

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

As another example we show that

$$(A_1 \cup A_2 \cup \cdots \cup A_n)^c = A_1{}^c \cap A_2{}^c \cap \cdots \cap A_n{}^c \qquad (1.2.3)$$

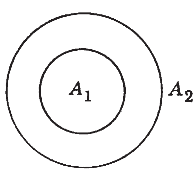


FIGURE 1.2.3 $A_1 \subset A_2$.

The steps are as follows.

$$\begin{aligned}
x \in (A_1 \cup \cdots \cup A_n)^c &\Leftrightarrow x \notin A_1 \cup \cdots \cup A_n \\
&\Leftrightarrow \text{it is not the case that } x \text{ belongs to at least one of the } A_i \\
&\Leftrightarrow x \in \text{none of the } A_i \\
&\Leftrightarrow x \in A_i^c \qquad \text{for all } i \\
&\Leftrightarrow x \in A_1^c \cap \cdots \cap A_n^c
\end{aligned}$$

An identical argument shows that

$$\left(\bigcup_{i=1}^{\infty} A_i\right)^c = \bigcap_{i=1}^{\infty} A_i^c \tag{1.2.4}$$

and similarly

$$\left(\bigcap_{i=1}^{n} A_i\right)^c = \bigcup_{i=1}^{n} A_i^c \qquad \text{i.e. } (A_1 \cap \cdots \cap A_n)^c = A_1^c \cup \cdots \cup A_n^c \tag{1.2.5}$$

Also

$$\left(\bigcap_{i=1}^{\infty} A_i\right)^c = \bigcup_{i=1}^{\infty} A_i^c \tag{1.2.6}$$

The identities (1.2.3)–(1.2.6) are called the *DeMorgan laws.*

In many ways the algebra of events differs from the algebra of real numbers, as some of the identities below indicate.

$$\begin{aligned}
A \cup A &= A & A \cup A^c &= \Omega \\
A \cap A &= A & A \cap A^c &= \varnothing \\
A \cap \Omega &= A & A \cup \varnothing &= A \\
A \cup \Omega &= \Omega & A \cap \varnothing &= \varnothing
\end{aligned} \tag{1.2.7}$$

Another method of verifying relations among events involves algebraic manipulation, using the identities already derived. Four examples are given below; in working out the identities, it may be helpful to write $A \cup B$ as $A + B$ and $A \cap B$ as AB.

$$1. \quad A \cup (A \cap B) = A \tag{1.2.8}$$

PROOF.

$$A + AB = A\Omega + AB = A(\Omega + B) = A\Omega = A$$

$$2. \quad (A \cup B) \cap (A \cup C) = A \cup (B \cap C) \tag{1.2.9}$$

PROOF.

$$\begin{aligned}(A+B)(A+C) &= (A+B)A + (A+B)C \\ &= AA + AB + AC + BC \qquad \text{(note } AB = BA\text{)} \\ &= A(\Omega + B + C) + BC \\ &= A\Omega + BC \\ &= A + BC\end{aligned}$$

$$3. \quad A \cup [(A \cap B)^c] = \Omega \tag{1.2.10}$$

PROOF.

$$A + (AB)^c = A + A^c + B^c = \Omega + B^c = \Omega$$

$$4. \quad (A \cap B^c) \cup (A \cap B) \cup (A^c \cap B) = A \cup B \tag{1.2.11}$$

PROOF.

$$\begin{aligned}AB^c + AB + A^cB &= AB^c + AB + AB + A^cB \qquad \text{[see (1.2.7)]} \\ &= A(B^c + B) + (A + A^c)B \\ &= A\Omega + \Omega B \\ &= A + B\end{aligned}$$

(see Figure 1.2.4).

As another example, let Ω be the set of nonnegative real numbers. Let

$$A_n = \left[0, 1 - \frac{1}{n}\right] = \left\{x \in \Omega: 0 \le x \le 1 - \frac{1}{n}\right\} \qquad n = 1, 2, \ldots$$

(This will be another common way of describing an event. It is to be read: "A_n is the set consisting of those points x in Ω such that $0 \le x \le 1 - 1/n$." If there is no confusion about what space Ω we are considering, we shall simply write $A_n = \{x: 0 \le x \le 1 - 1/n\}$.) Then

$$\bigcup_{n=1}^{\infty} A_n = [0, 1) = \{x: 0 \le x < 1\}$$

$$\bigcap_{n=1}^{\infty} A_n = \{0\}$$

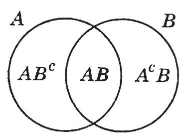


FIGURE 1.2.4 Venn Diagram Illustrating $(A \cap B^c) \cup (A \cap B) \cup (A^c \cap B) = A \cup B$.

As an illustration of the DeMorgan laws,

$$\left(\bigcup_{n=1}^{\infty} A_n\right)^c = [0, 1)^c = [1, \infty) = \{x\colon x \geq 1\}$$

$$\bigcap_{n=1}^{\infty} A_n{}^c = \bigcap_{n=1}^{\infty} \left(1 - \frac{1}{n}, \infty\right) = [1, \infty)$$

(Notice that $x > 1 - 1/n$ for all $n = 1, 2, \ldots \Leftrightarrow x \geq 1$.) Also

$$\left(\bigcap_{n=1}^{\infty} A_n\right)^c = \{0\}^c = (0, \infty) = \{x\colon x > 0\}$$

$$\bigcup_{n=1}^{\infty} A_n{}^c = \bigcup_{n=1}^{\infty} \left(1 - \frac{1}{n}, \infty\right) = (0, \infty)$$

PROBLEMS

1. An experiment involves choosing an integer N between 0 and 9 (the sample space consists of the integers from 0 to 9, inclusive). Let $A = \{N \leq 5\}$, $B = \{3 \leq N \leq 7\}$, $C = \{N \text{ is even and } N > 0\}$. List the points that belong to the following events.

$$A \cap B \cap C, \qquad A \cup (B \cap C^c), \qquad (A \cup B) \cap C^c, \qquad (A \cap B) \cap [(A \cup C)^c]$$

2. Let A, B, and C be *arbitrary* events in the same sample space. Let D_1 be the event that at least two of the events A, B, C occur; that is, D_1 is the set of points common to at least two of the sets A, B, C.
Let $D_2 = \{$exactly two of the events A, B, C occur$\}$
$D_3 = \{$at least one of the events A, B, C occur$\}$
$D_4 = \{$exactly one of the events A, B, C occur$\}$
$D_5 = \{$not more than two of the events A, B, C occur$\}$
Each of the events D_1 through D_5 can be expressed in terms of A, B, and C by using unions, intersections, and complements. For example, $D_3 = A \cup B \cup C$. Find suitable expressions for D_1, D_2, D_4, and D_5.

3. A public opinion poll (circa 1850) consisted of the following three questions:
(a) Are you a registered Whig?
(b) Do you approve of President Fillmore's performance in office?
(c) Do you favor the Electoral College system?
A group of 1000 people is polled. Assume that the answer to each question must be either "yes" or "no." It is found that:
550 people answer "yes" to the third question and 450 answer "no."
325 people answer "yes" exactly twice; that is, their responses contain two "yeses" and one "no."

100 people answer "yes" to all three questions.
125 registered Whigs approve of Fillmore's performance.

How many of those who favor the Electoral College system do not approve of Fillmore's performance, and in addition are not registered Whigs? HINT: Draw a Venn diagram.

4. If A and B are events in a sample space, define $A - B$ as the set of points which belong to A but not to B; that is, $A - B = A \cap B^c$. Establish the following.
(a) $A \cap (B - C) = (A \cap B) - (A \cap C)$
(b) $A - (B \cup C) = (A - B) - C$
Is is true that $(A - B) \cup C = (A \cup C) - B$?

5. Let Ω be the reals. Establish the following.

$$(a, b) = \bigcup_{n=1}^{\infty}\left(a, b - \frac{1}{n}\right] = \bigcup_{n=1}^{\infty}\left[a + \frac{1}{n}, b\right)$$

$$[a, b] = \bigcap_{n=1}^{\infty}\left[a, b + \frac{1}{n}\right) = \bigcap_{n=1}^{\infty}\left(a - \frac{1}{n}, b\right]$$

6. If A and B are disjoint events, are A^c and B^c disjoint? Are $A \cap C$ and $B \cap C$ disjoint? What about $A \cup C$ and $B \cup C$?

7. If $A_n \subset A_{n-1} \subset \cdots \subset A_1$, show that $\bigcap_{i=1}^{n} A_i = A_n$, $\bigcup_{i=1}^{n} A_i = A_1$.

8. Suppose that $A_1, A_2, \ldots$ is a sequence of subsets of Ω, and we know that for each n, $\bigcap_{i=1}^{n} A_i$ is not empty. Is it true that $\bigcap_{i=1}^{\infty} A_i$ is not empty? (A related question about real numbers: if, for each n, we have $\sum_{i=1}^{n} a_i < b$, is it true that $\sum_{i=1}^{\infty} a_i < b$?)

9. If $A, B_1, B_2, \ldots$ are arbitrary events, show that

$$A \cap \left(\bigcup_i B_i\right) = \bigcup_i (A \cap B_i)$$

This is the distributive law with infinitely many factors.

1.3 PROBABILITY

We now consider the assignment of probabilities to events. A technical complication arises here. It may not always be possible to regard all subsets of Ω as events. We may discard or fail to measure some of the information in the outcome corresponding to the point $\omega \in \Omega$, so that for a given subset A of Ω, it may not be possible to give a yes or no answer to the question "Is $\omega \in A$?" For example, if the experiment involves tossing a coin five times, we may record the results of only the first three tosses, so that $A = \{$at least four heads$\}$ will not be "measurable"; that is, membership of $\omega \in A$ cannot be determined from the given information about ω.

In a given problem there will be a particular class of subsets of Ω called the "class of events." For reasons of mathematical consistency, we require that the event class $\mathscr{F}$ form a *sigma field*, which is a collection of subsets of Ω satisfying the following three requirements.

$$\Omega \in \mathscr{F} \tag{1.3.1}$$

$$A_1, A_2, \ldots \in \mathscr{F} \quad \text{implies} \quad \bigcup_{n=1}^{\infty} A_n \in \mathscr{F} \tag{1.3.2}$$

That is, $\mathscr{F}$ is closed under finite or countable union.

$$A \in \mathscr{F} \quad \text{implies} \quad A^c \in \mathscr{F} \tag{1.3.3}$$

That is, $\mathscr{F}$ is closed under complementation.

Notice that if $A_1, A_2, \ldots \in \mathscr{F}$, then $A_1^c, A_2^c, \ldots \in \mathscr{F}$ by (1.3.3); hence $\bigcup_{n=1}^{\infty} A_n^c \in \mathscr{F}$ by (1.3.2). By the DeMorgan laws, $\bigcap_{n=1}^{\infty} A_n = (\bigcup_{n=1}^{\infty} A_n^c)^c$; hence, by (1.3.3), $\bigcap_{n=1}^{\infty} A_n \in \mathscr{F}$. Thus $\mathscr{F}$ is closed under finite or countable intersection. Also, by (1.3.1) and (1.3.3), the empty set $\varnothing$ belongs to $\mathscr{F}$.

Thus, for example, if the question "Did A_n occur?" has a definite answer for $n = 1, 2, \ldots$, so do the questions "Did at least one of the A_n occur?" and "Did all the A_n occur?"

Note also that if we apply the algebraic operations of Section 1.2 to sets in $\mathscr{F}$, the new sets we obtain still belong to $\mathscr{F}$.

In many cases we shall be able to take $\mathscr{F}$ = the collection of *all* subsets of Ω, so that every subset of Ω is an event. Problems in which $\mathscr{F}$ cannot be chosen in this way generally arise in uncountably infinite sample spaces; for example, Ω = the reals. We shall return to this subject in Chapter 2.

We are now ready to talk about the assignment of probabilities to events. If $A \in \mathscr{F}$, the probability $P(A)$ should somehow reflect the long-run relative frequency of A in a large number of independent repetitions of the experiment. Thus $P(A)$ should be a number between 0 and 1, and $P(\Omega)$ should be 1.

Now if A and B are disjoint events, the number of occurrences of $A \cup B$ in n performances of the experiment is obtained by adding the number of occurrences of A to the number of occurrences of B. Thus we should have

$$P(A \cup B) = P(A) + P(B) \qquad \text{if } A \text{ and } B \text{ are disjoint}$$

and, similarly,

$$P(A_1 \cup \cdots \cup A_n) = \sum_{i=1}^{n} P(A_i) \qquad \text{if } A_1, \ldots, A_n \text{ are disjoint}$$

For mathematical convenience we require that

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

when we have a *countably infinite* family of disjoint events $A_1, A_2, \ldots$.

The assumption of countable rather than simply finite additivity has not been convincingly justified physically or philosophically; however, it leads to a much richer mathematical theory.

A function that assigns a number $P(A)$ to each set A in the sigma field $\mathscr{F}$ is called a *probability measure* on $\mathscr{F}$, provided that the following conditions are satisfied.

$$P(A) \geq 0 \qquad \text{for every } A \in \mathscr{F} \tag{1.3.4}$$

$$P(\Omega) = 1 \tag{1.3.5}$$

If $A_1, A_2, \ldots$ are disjoint sets in $\mathscr{F}$, then

$$P(A_1 \cup A_2 \cup \cdots) = P(A_1) + P(A_2) + \cdots \tag{1.3.6}$$

We may now give the underlying mathematical framework for probability theory.

DEFINITION. A *probability space* is a triple $(\Omega, \mathscr{F}, P)$, where Ω is a set, $\mathscr{F}$ a sigma field of subsets of Ω, and P a probability measure on $\mathscr{F}$.

We shall not, at this point, embark on a general study of probability measures. However, we shall establish four facts from the definition. (All sets in the arguments to follow are assumed to belong to $\mathscr{F}$.)

$$1. \quad P(\varnothing) = 0 \tag{1.3.7}$$

PROOF. $A \cup \varnothing = A$; hence $P(A \cup \varnothing) = P(A)$. But A and $\varnothing$ are disjoint and so $P(A \cup \varnothing) = P(A) + P(\varnothing)$. Thus $P(A) = P(A) + P(\varnothing)$; consequently $P(\varnothing) = 0$.

$$2. \quad P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{1.3.8}$$

PROOF. $A = (A \cap B) \cup (A \cap B^c)$, and these sets are disjoint (see Figure 1.2.4). Thus $P(A) = P(A \cap B) + P(A \cap B^c)$. Similarly $P(B) = P(A \cap B) + P(A^c \cap B)$. Thus $P(A) + P(B) - P(A \cap B) = P(A \cap B) + P(A \cap B^c) + P(A^c \cap B) = P(A \cup B)$. Intuitively, if we add the outcomes in A to those in B, we have counted those in $A \cap B$ twice; subtracting the outcomes in $A \cap B$ yields the outcomes in $A \cup B$.

3. If $B \subset A$, then $P(B) \leq P(A)$; in fact,

$$P(A - B) = P(A) - P(B) \tag{1.3.9}$$

where $A - B$ is the set of points that belong to A but not to B.

PROOF. $P(A) = P(B) + P(A - B)$, since $B \subset A$ (see Figure 1.3.1), and the result follows because $P(A - B) \geq 0$. Intuitively, if the occurrence of B

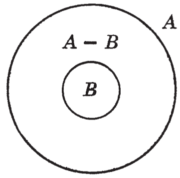


FIGURE 1.3.1

always implies the occurrence of A, A must occur at least as often as B in any sequence of performances of the experiment.

$$4. \quad P(A_1 \cup A_2 \cup \cdots) \le P(A_1) + P(A_2) + \cdots \tag{1.3.10}$$

That is, the probability that at least one of a finite or countably infinite collection of events will occur is less than or equal to the sum of the probabilities; note that, for the case of two events, this follows from $P(A \cup B) = P(A) + P(B) - P(A \cap B) \le P(A) + P(B)$.

PROOF. We make use of the fact that any union may be written as a disjoint union, as follows.

$$A_1 \cup A_2 \cup \cdots = A_1 \cup (A_1{}^c \cap A_2) \cup (A_1{}^c \cap A_2{}^c \cap A_3) \cup \cdots \cup (A_1{}^c \cap A_2{}^c \cap \cdots \cap A_{n-1}^c \cap A_n) \cup \cdots \tag{1.3.11}$$

To see this, observe that if x belongs to the set on the right then $x \in A_1{}^c \cap \cdots \cap A_{n-1}^c \cap A_n$ for some n; hence $x \in A_n$. Thus x belongs to the set on the left. Conversely, if x belongs to the set on the left, then $x \in A_n$ for some n. Let n_0 be the smallest such n. Then $x \in A_1{}^c \cap \cdots \cap A_{n_0-1}^c \cap A_{n_0}$, and so x belongs to the set on the right. Thus

$$P(A_1 \cup A_2 \cup \cdots) = \sum_{n=1}^{\infty} P(A_1{}^c \cap \cdots \cap A_{n-1}^c \cap A_n) \le \sum_{n=1}^{\infty} P(A_n)$$

using (1.3.9); notice that

$$A_1{}^c \cap \cdots \cap A_{n-1}^c \cap A_n \subset A_n.$$

REMARKS. The basic difficulty with the classical and frequency definitions of probability is that their approach is to try somehow to *prove* mathematically that, for example, the probability of picking a heart from a perfectly shuffled deck is 1/4, or that the probability of an unbiased coin coming up heads is 1/2. This cannot be done. All we can say is that if a card is picked at random and then replaced, and the process is repeated over and over again, the result that the ratio of hearts to total number of drawings will be close to 1/4 is in accord with our intuition and our physical experience. For this reason we should *assign* a probability 1/4 to the event of obtaining a heart, and similarly we should *assign* a probability 1/52 to each possible outcome of the experiment. The only reason for doing this is that the consequences agree with our experience. If you decide that some mysterious factor caused the ace of spades to be more likely than any other card, you could incorporate this factor by assigning a higher probability to the ace of spades. The mathematical development of the theory would not be affected; however, the conclusions you might draw from this assumption would be at variance with experimental results.

One can never really use mathematics to *prove* a specific physical fact. For example, we cannot prove mathematically that there is a physical quantity called "force." What we can do is postulate a mathematical entity called "force" that satisfies a certain differential equation. We can build up a collection of mathematical results that, when interpreted properly, provide a reasonable description of certain physical phenomena (reasonable until another mathematical theory is constructed that provides a better description). Similarly, in probability theory we are faced with situations in which our intuition or some physical experiments we have carried out suggest certain results. Intuition and experience lead us to an *assignment* of probabilities to events. As far as the mathematics is concerned, any assignment of probabilities will do, subject to the rules of mathematical consistency. However, our hope is to develop mathematical results that, when interpreted and related to physical experience, will help to make precise such notions as "the ratio of the number of heads to the total number of observations in a very large number of independent tosses of an unbiased coin is very likely to be very close to 1/2."

We emphasize that the insights gained by the early workers in probability are not to be discarded, but instead cast in a more precise form.

PROBLEMS

1. Write down some examples of sigma fields other than the collection of all subsets of a given set Ω.

2. Give an example to show that $P(A - B)$ need not equal $P(A) - P(B)$ if B is not a subset of A.

1.4 COMBINATORIAL PROBLEMS

We consider a class of problems in which the assignment of probabilities can be made in a natural way.

Let Ω be a *finite* or *countably infinite* set, and let $\mathscr{F}$ consist of all subsets of Ω.

For each point $\omega_i \in \Omega$, $i = 1, 2, \ldots$, assign a nonnegative number p_i, with $\sum_i p_i = 1$. If A is any subset of Ω, let $P(A) = \sum_{\omega_i \in A} p_i$. Then it may be verified that P is a probability measure; $P\{\omega_i\} = p_i$, and *the probability of any event A is found by adding the probabilities of the points of A*. An $(\Omega, \mathscr{F}, P)$ of this type is called a *discrete probability space*.

▶ **Example 1.** Throw a (biased) coin twice (see Figure 1.4.1).

Let $E_1 = \{\text{at least one head}\}$. Then

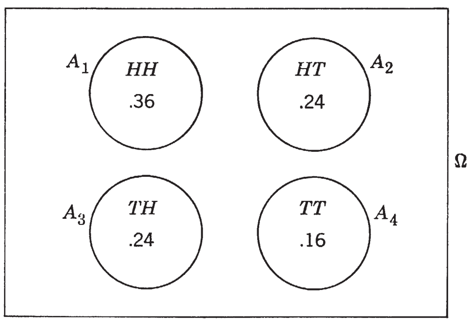
$$E_1 = A_1 \cup A_2 \cup A_3$$

Hence

$$\begin{aligned} P(E_1) &= P(A_1) + P(A_2) + P(A_3) \\ &= .36 + .24 + .24 = .84 \end{aligned}$$

Let $E_2 = \{\text{tail on first toss}\}$; then

$$E_2 = A_3 \cup A_4$$



$$A_1 = \{HH\}, \qquad A_2 = \{HT\}$$
$$A_3 = \{TH\}, \qquad A_4 = \{TT\}$$
$$\begin{aligned} \text{Assign } P(A_1) &= .36 \\ P(A_2) = P(A_3) &= .24 \\ P(A_4) &= .16 \end{aligned}$$

FIGURE 1.4.1 Coin-Tossing Problem.

and

$$P(E_2) = P(A_3) + P(A_4) = .4 \blacktriangleleft$$

In the special case when $\Omega = \{\omega_1, \ldots, \omega_n\}$ and $p_i = 1/n$, $i = 1, 2, \ldots, n$, we have

$$P(A) = \frac{\text{number of points of } A}{\text{total number of points in } \Omega} = \frac{\text{favorable outcomes}}{\text{total outcomes}}$$

corresponding to the classical definition of probability.

Thus, in this case, finding $P(A)$ simply involves counting the number of outcomes favorable to A. When n is large, counting by hand may not be feasible; combinatorial analysis is simply a method of counting that can often be used to avoid writing down the entire list of favorable outcomes.

There is only one basic idea in combinatorial analysis, and that is the following. Suppose that a symbol is selected from the set $\{a_1, \ldots, a_n\}$; if a_i is chosen, a symbol is selected from the set $\{b_{i1}, \ldots, b_{im}\}$. Each pair of selections (a_i, b_{ij}) is assumed to determine a "result" $f(i, j)$. If all results are distinct, the number of possible results is nm, since there is a one-to-one correspondence between results and pairs of integers (i, j), $i = 1, \ldots, n$, $j = 1, \ldots, m$.

If, after the symbol b_{ij} is chosen, a symbol is selected from the set $\{c_{ij1}, c_{ij2}, \ldots, c_{ijp}\}$, and each triple (a_i, b_{ij}, c_{ijk}) determines a distinct result $f(i, j, k)$, the number of possible results is nmp. Analogous statements may be made for any finite sequence of selections.

Certain standard selections occur frequently, and it is convenient to classify them.

Let $a_1, \ldots, a_n$ be distinct symbols.

Ordered samples of size r, with replacement

The number of ordered sequences $(a_{i_1}, \ldots, a_{i_r})$, where the a_{i_k} belong to $\{a_1, \ldots, a_n\}$, is $n \times n \times \cdots \times n$ (r times), or

$$n^r \tag{1.4.1}$$

(The term "with replacement" refers to the fact that if the symbol a_{i_k} is selected at step k it may be selected again at any future time.)

For example, the number of possible outcomes if three dice are thrown is $6 \times 6 \times 6 = 216$.

Ordered Samples of Size r, without Replacement

The number of ordered sequences $(a_{i_1}, \ldots, a_{i_r})$, where the a_{i_k} belong to $\{a_1, \ldots, a_n\}$, but repetition is not allowed (i.e., no a_i can appear more than once in the sequence), is

$$n(n-1)\cdots(n-r+1) = \frac{n!}{(n-r)!}, \qquad r = 1, 2, \ldots, n \quad (1.4.2)$$

(The first symbol may be chosen in n ways, and the second in $n-1$ ways, since the first symbol may not be used again, and so on.) The above number is sometimes called the number of *permutations of r objects out of n*, written $(n)_r$.

For example, the number of 3-digit numbers that can be formed from $1, 2, \ldots, 9$, if no digit can be repeated, is $9(8)(7) = 504$.

Unordered Samples of Size r, without Replacement

The number of unordered sets $\{a_{i_1}, \ldots, a_{i_r}\}$, where the a_{i_k}, $k = 1, \ldots, r$, are distinct elements of $\{a_1, \ldots, a_n\}$ (i.e., the number of ways of selecting r distinct objects out of n), if order does not count, is

$$\binom{n}{r} = \frac{n!}{r!\,(n-r)!} \quad (1.4.3)$$

To see this, consider the following process.

(a) Select r distinct objects out of n without regard to order; this can be done in $\binom{n}{r}$ ways, where $\binom{n}{r}$ is to be determined.

(b) For each set selected in (a), say $\{a_{i_1}, \ldots, a_{i_r}\}$, select an ordering of $a_{i_1}, \ldots, a_{i_r}$. This can be done in $(r)_r = r!$ ways (see Figure 1.4.2 for $n = 3$, $r = 2$).

The result of performing (a) and (b) is a permutation of r objects out of n; hence

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

or

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

We define $\binom{n}{0}$ to be $n!/0!\,n! = 1$, to make the formula for $\binom{n}{r}$ valid for $r = 0, 1, \ldots, n$. Notice that $\binom{n}{k} = \binom{n}{n-k}$.

$\binom{n}{r}$ is sometimes called the number of *combinations of r objects out of n*.

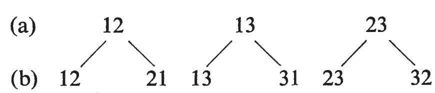


FIGURE 1.4.2 Determination of $\binom{n}{r}$.

Unordered Samples of Size r, with Replacement

We wish to find the number of unordered sets $\{a_{i_1}, \ldots, a_{i_r}\}$, where the a_{i_k} belong to $\{a_1, \ldots, a_n\}$ and repetition is allowed. As an example, let $n = 3$ and $r = 3$. Let the symbols be 1, 2, and 3. List all arrangements in a column so that a precedes b if and only if a, read as an ordinary 3-digit number, is $<b$. In an adjacent column list a new set of sequences formed from the old by adding 0 to the first digit, 1 to the second digit, and 2 to the third digit.

111	$123 = (1 + 0, 1 + 1, 1 + 2)$
112	$124 = (1 + 0, 1 + 1, 2 + 2)$
113	125
122	134
123	135
133	145
222	234
223	235
233	245
333	345

In the first column we have unordered samples of size 3 (out of 3), with replacement. In the second column we have unordered samples of size 3 (out of 5), without replacement. In this way we can set up a one-to-one correspondence between unordered samples of size r (out of n) with replacement, and unordered samples of size r (out of $n + r - 1$) without replacement. Thus the number of such samples is

$$\binom{n + r - 1}{r} \tag{1.4.4}$$

An alternative way of looking at unordered samples with replacement is to count all sequences $(a_{i_1}, \ldots, a_{i_r})$, each $a_{i_k} \in \{a_1, \ldots, a_n\}$, subject to the constraint that sequences having the same *occupancy numbers* $r_k =$ the number of occurrences of a_k, $k = 1, 2, \ldots, n$, are identified. The r_k are nonnegative integers satisfying $r_1 + r_2 + \cdots r_n = r$; hence we must count the number of nonnegative integer solutions $(r_1, \ldots, r_n)$ of the equation $r_1 + \cdots + r_n = r$. This may be done combinatorally as follows.

Consider an arrangement of r stars and $n - 1$ bars, as shown in Figure

$\vert * \vert \vert *** \vert$

$r_1 = 1, r_2 = 0, r_3 = 3$

(sample $= a_1a_3a_3a_3$)

$\vert * \vert * \vert * * \vert$

$r_1 = 1, r_2 = 1, r_3 = 2$

(sample $= a_1a_2a_3a_3$)

FIGURE 1.4.3 Counting Unordered Samples with Replacement.

1.4.3 for $n = 3$, $r = 4$ (the thicker bars at the sides are fixed). Each arrangement corresponds to a solution of $r_1 + \cdots + r_n = r$. The number of arrangements is the number of ways of selecting r positions out of $n + r - 1$ for the stars to occur (or $n - 1$ positions for the bars); that is, $\binom{n+r-1}{r}$. For $n = 3$, $r = 4$, there are 15 solutions.

r_1	r_2	r_3	*Sample*
0	0	4	$a_3a_3a_3a_3$
0	4	0	$a_2a_2a_2a_2$
4	0	0	$a_1a_1a_1a_1$
0	1	3	$a_2a_3a_3a_3$
0	2	2	$a_2a_2a_3a_3$
0	3	1	$a_2a_2a_2a_3$
1	0	3	$a_1a_3a_3a_3$
2	0	2	$a_1a_1a_3a_3$
3	0	1	$a_1a_1a_1a_3$
1	3	0	$a_1a_2a_2a_2$
2	2	0	$a_1a_1a_2a_2$
3	1	0	$a_1a_1a_1a_2$
2	1	1	$a_1a_1a_2a_3$
1	2	1	$a_1a_2a_2a_3$
1	1	2	$a_1a_2a_3a_3$

▶ **Example 2.** Find the probability of obtaining four of a kind in an ordinary five-card poker hand.

There are $\binom{52}{5}$ distinct poker hands (without regard to order), and so we may take Ω to have $\binom{52}{5}$ points. To obtain the number of hands in which there are four of a kind:

(a) Choose the face value to appear four times (13 choices: A, K, Q, . . . , 2)

(b) Choose the fifth card (48 ways).

Thus $p = (13)(48)/\binom{52}{5}$. Figure 1.4.4 indicates the selection process.

NOTE. The problem may also be done using ordered samples. The number of ordered poker hands is $(52)(51)(50)(49)(48) = (52)_5$ (the drawing is without replacement). The number of ordered poker hands having four of a kind is $(13)(48)\ 5!$, so that $p = (13)(48)(5!)/(52)_5 = (13)(48)/\binom{52}{5}$ as before. Here we may take the space Ω' to have $(52)_5$ points; each point of Ω corresponds to $5!$ points of Ω'. ◀

▶ **Example 3.** Three balls are dropped into three boxes. Find the probability that exactly one box will be empty.

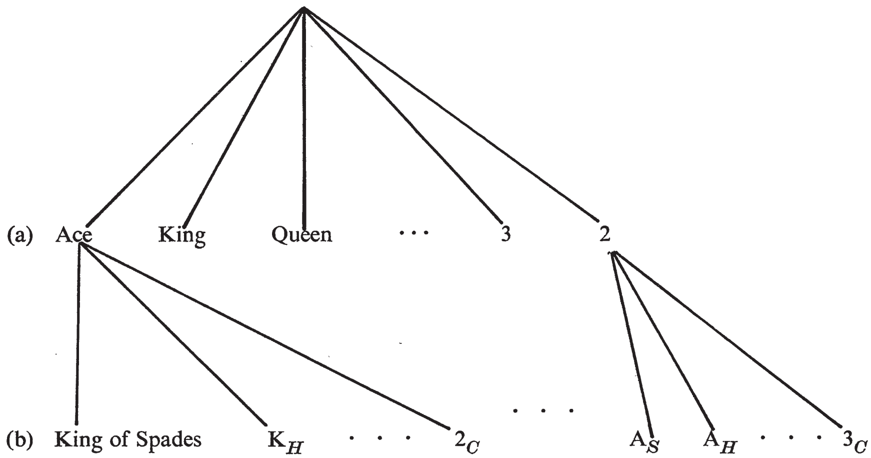


FIGURE 1.4.4 Counting Process for Selecting Four of a Kind. There is a one-to-one correspondence between paths in the diagram and favorable outcomes.

In problems of dropping r balls into n boxes, we may regard the boxes as (distinct) symbols $a_1, \ldots, a_n$; each toss of a ball corresponds to the selection of a box. Thus the sequence $a_2a_1a_1a_3$ corresponds to the first ball into box 2, the second and third into box 1, and the fourth into box 3.

In general, an arrangement of r balls in n boxes corresponds to a sample of size r from the symbols $a_1, \ldots, a_n$. If we require that the sampling be with replacement, this means that a given box can contain any number of balls. Sampling without replacement means that a given box cannot contain more than one ball. If we consider ordered samples we are saying that the balls are distinguishable. For example, a_3a_7 (ball 1 into box 3, ball 2 into box 7) is different from a_7a_3 (ball 1 into box 7, ball 2 into box 3); in other words, we may regard the balls as being numbered $1, 2, \ldots, r$. Unordered sampling corresponds to indistinguishable balls.

If there is no restriction on the number of balls in a given box, the total number of arrangements, taking into account the order in which the balls are tossed (i.e., regarding the balls as distinct), is the number of ordered samples of size r (from $\{a_1, \ldots, a_n\}$) with replacement, or n^r. If the boxes are energy levels in physics and the balls are particles, the *Maxwell-Boltzmann* assumption is that all n^r arrangements are equally likely.

If there can be at most one ball in a given box, the number of (ordered) arrangements is $(n)_r$. If the order in which the balls are tossed is neglected, we are simply choosing r boxes out of n to be occupied; the *Fermi-Dirac* assumption takes the $\binom{n}{r}$ possible selections of boxes (or energy levels) as equally likely.

We might also mention the *Bose-Einstein* assumption; here a box may contain an unlimited number of balls, but the balls are indistinguishable; that is, the order in which the balls are tossed is neglected, so that, for example, $a_2a_1a_1a_3$ is identified with $a_1a_3a_1a_2$. Thus the number of arrangements counted in this scheme is the number of unordered samples of size r with replacement, or $\binom{n+r-1}{r}$. The Bose-Einstein assumption takes all these arrangements as equally likely.

To return to the original problem, we have boxes a_1, a_2, and a_3 and sequences of length 3 (three balls are tossed). We take all $3^3 = 27$ ordered samples with replacement as equally likely. (We shall see that this model—ordered sampling with replacement—corresponds to the tossing of the balls independently; this idea will be developed in the next section.) Now

$$\begin{aligned} P\{\text{exactly 1 box empty}\} &= P\{\text{box 1 empty, boxes 2 and 3 occupied}\} \\ &\quad + P\{\text{box 2 empty, boxes 1 and 3 occupied}\} \\ &\quad + P\{\text{box 3 empty, boxes 1 and 2 occupied}\} \end{aligned}$$

Furthermore

$$P\{\text{box 1 empty, boxes 2 and 3 occupied}\} = P\{a_1 \text{ does not occur in the sequence } a_{i_1}a_{i_2}a_{i_3}, \text{ but } a_2 \text{ and } a_3 \text{ both occur}\}$$

If a_1 does not occur, either a_2 or a_3 must occur twice, and the other symbol once. We may choose the symbol that is to occur twice in two ways; the symbol that occurs once is then determined. If, say, a_3 occurs twice and a_2 once, the position of a_2 may be any of three possibilities; the position of the two a_3's is then determined. Thus the probability that box 1 will be empty and boxes 2 and 3 occupied is $2(3)/27 = 6/27$ (in fact the six favorable outcomes are $a_2a_2a_3$, $a_2a_3a_2$, $a_3a_2a_2$, $a_3a_3a_2$, $a_3a_2a_3$, and $a_2a_3a_3$).

Thus the probability that exactly one box will be empty is, by symmetry, $3(6)/27 = 2/3$. ◀

▶ **Example 4.** In a 13-card bridge hand the probability that the hand will contain the A K Q J 10 of spades is $\binom{47}{8}/\binom{52}{13}$. (The A K Q J 10 of spades must be chosen, and afterward eight cards must be selected out of 47 that remain after the five top spades have been removed.)

Now let us find the probability of obtaining the A K Q J 10 of at least one suit. Thus, if A_S is the event that the A K Q J 10 of spades is obtained, and similarly for A_H, A_D, and A_C (hearts, diamonds, and clubs), we are looking for

$$P(A_S \cup A_H \cup A_D \cup A_C)$$

The sets are not disjoint, so that we cannot simply add probabilities. It is possible to obtain, for example, the A K Q J 10 of both spades and hearts in a 13-card hand, and this probability is easy to compute:

$$P(\mathrm{A_S} \cap \mathrm{A_H}) = \frac{\binom{42}{3}}{\binom{52}{13}}$$

What we need here is a way of expressing $P(\mathrm{A_S} \cup \mathrm{A_H} \cup \mathrm{A_D} \cup \mathrm{A_C})$ in terms of the individual terms $P(\mathrm{A_S})$ etc., and the intersections $P(\mathrm{A_S} \cap \mathrm{A_H})$ etc. We know that

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

If we have three events, then

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

The general pattern is now clear and may be verified by induction.

$$\begin{aligned} P(A_1 \cup \cdots \cup A_n) &= \sum_i P(A_i) - \sum_{i<j} P(A_i \cap A_j) \\ &+ \sum_{i<j<k} P(A_i \cap A_j \cap A_k) - \cdots + (-1)^{n-1} P(A_1 \cap \cdots \cap A_n) \end{aligned} \quad (1.4.5)$$

In the present problem the intersections taken three or four at a time are empty; hence, by symmetry,

$$\begin{aligned} P(\mathrm{A_S} \cup \mathrm{A_H} \cup \mathrm{A_D} \cup \mathrm{A_C}) &= 4P(\mathrm{A_S}) - 6P(\mathrm{A_S} \cap \mathrm{A_H}) \\ &= \frac{4\binom{47}{8} - 6\binom{42}{3}}{\binom{52}{13}} \end{aligned}$$

It is illuminating to consider an *incorrect* approach to this problem. Suppose that we first pick a suit (four choices); we then select the A K Q J 10 of that suit. The remaining eight cards can be anything (if they include the A K Q J 10 of another suit, the condition that at least one A K Q J 10 of the same suit be obtained will still be satisfied). Thus we have $\binom{47}{8}$ choices, so that the desired probability is $4\binom{47}{8}/\binom{52}{13}$.

The above procedure illustrates *multiple counting*, the nemesis of the combinatorial analyst (see Figure 1.4.5).

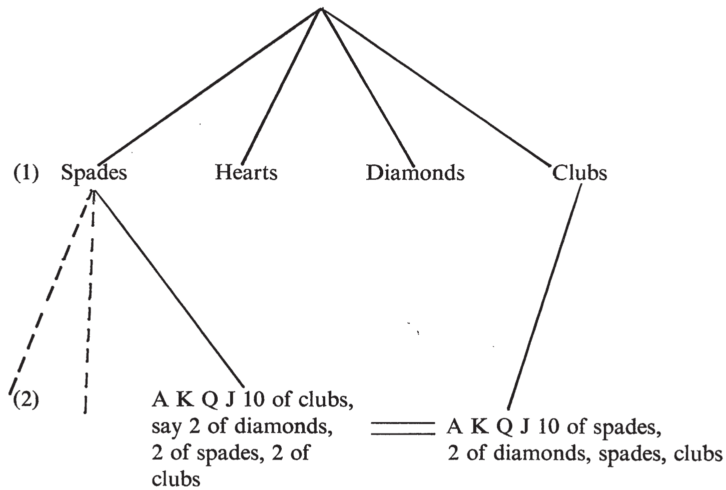


FIGURE 1.4.5 Multiple Counting.

In writing $p = 4\binom{47}{8}/\binom{52}{13}$ we are saying that there is a one-to-one correspondence between paths in Figure 1.4.5 and favorable outcomes. But this is not the case, since the two paths indicated in the diagram both lead to the same result, namely, A K Q J 10 of spades, A K Q J 10 of clubs, 2 of diamonds, 2 of spades, 2 of clubs. In fact there are $6\binom{42}{3}$ such duplications. For we can pick the two suits in $\binom{4}{2} = 6$ ways; then, after taking A K Q J 10 of each suit, we select the remaining three cards in $\binom{42}{3}$ ways. If we subtract the number of duplications, $6\binom{42}{3}$, from the original count, $4\binom{47}{8}$, we obtain the correct result.

To rephrase: the counting process we have proposed counts the number of paths in the above diagram, that is, the number of choices at Step 1 times the number of choices at Step 2. However, the paths do not in general lead to distinct "results," namely, distinct bridge hands. ◀

PROBLEMS

1. If a 3-digit number (000 to 999) is chosen at random, find the probability that exactly 1 digit will be >5.
2. Find the probability that a five-card poker hand will be:
   (a) A straight (five cards in sequence regardless of suit; ace may be high but not low).

(b) Three of a kind (three cards of the same face value x, plus two cards with face values y and z, with x, y, z distinct).
(c) Two pairs (two cards of face value x, two of face value y, and one of face value z, with x, y, z distinct).

3. An urn contains 3 red, 8 yellow, and 13 green balls; another urn contains 5 red, 7 yellow, and 6 green balls. One ball is selected from each urn. Find the probability that both balls will be of the same color.

4. An experiment consists of drawing 10 cards from an ordinary 52-card pack.
(a) If the drawing is done with replacement, find the probability that no two cards will have the same face value.
(b) If the drawing is done without replacement, find the probability that at least 9 cards will be of the same suit.

5. An urn contains 10 balls numbered from 1 to 10. Five balls are drawn without replacement. Find the probability that the second largest of the five numbers drawn will be 8.

6. m men and w women seat themselves at random in $m + w$ seats arranged in a row. Find the probability that all the women will be adjacent.

7. If a box contains 75 good light bulbs and 25 defective bulbs and 15 bulbs are removed, find the probability that at least one will be defective.

8. Eight cards are drawn without replacement from an ordinary deck. Find the probability of obtaining exactly three aces or exactly three kings (or both).

9. (The game of rencontre). An urn contains n tickets numbered $1, 2, \ldots, n$. The tickets are shuffled thoroughly and then drawn one by one without replacement. If the ticket numbered r appears in the rth drawing, this is denoted as a *match* (French: *rencontre*). Show that the probability of at least one match is

$$1 - \frac{1}{2!} + \frac{1}{3!} - \cdots + \frac{(-1)^{n-1}}{n!} \to 1 - e^{-1} \quad \text{as} \quad n \to \infty$$

10. A "language" consists of three "words," $W_1 = a$, $W_2 = ba$, $W_3 = bb$. Let $N(k)$ be the number of "sentences" using exactly k letters (e.g., $N(1) = 1$ (i.e., a), $N(2) = 3$ (aa, ba, bb), $N(3) = 5$ (aaa, aba, abb, baa, bba); no space is allowed between words).
(a) Show that $N(k) = N(k - 1) + 2N(k - 2)$, $k = 2, 3, \ldots$ (define $N(0) = 1$).
(b) Show that the general solution to the second-order homogeneous linear difference equation (a) [with $N(0)$ and $N(1)$ specified], is $N(k) = A2^k + B(-1)^k$, where A and B are determined by $N(0)$ and $N(1)$. Evaluate A and B in the present case.

11. (The birthday problem) Assume that a person's birthday is equally likely to fall on any of the 365 days in a year (neglect leap years). If r people are selected, find the probability that all r birthdays will be different. Equivalently, if r balls are dropped into 365 boxes, we are looking for the probability that no box will contain more than one ball. It turns out that the probability is less than 1/2

for $r \geq 23$, so that in a class of 23 or more students the odds are that two or more people will have the same birthday.

12. Fourteen balls are dropped into six boxes. Find the number of arrangements (ordered samples of size 14 with replacement, from six symbols) whose occupancy numbers coincide with 4, 4, 2, 2, 2, 0 in some order (i.e., boxes i_1 and i_2 contain four balls, boxes i_3, i_4, and i_5 two balls, and box i_6 no balls, for some $i_1, \ldots, i_6$).

13. (a) Let Ω be a set with n elements. Show that there are 2^n subsets of Ω. For example, if $\Omega = \{1, 2, 3\}$, the subsets are $\varnothing$, $\{1\}$, $\{2\}$, $\{3\}$, $\{1, 2\}$, $\{1, 3\}$, $\{2, 3\}$, and $\{1, 2, 3\} = \Omega$; $2^3 = 8$ altogether.
(b) How many ways are there of selecting ordered pairs (A, B) of subsets of Ω such that $A \subset B$? For example, $A = \{1\}$, $B = \{1, 3\}$ gives such a pair, but $A = \{1, 2\}$, $B = \{1, 3\}$ does not.

14. Let Ω be a finite set. A *partition* of Ω is an (unordered) set $\{A_1, \ldots, A_n\}$, where the A_i are nonempty subsets whose union is Ω. For example, if $\Omega = \{1, 2, 3\}$, there are five partitions.

$$\begin{aligned}
A_1 &= \{1, 2, 3\} \\
A_1 &= \{1, 2\}, A_2 = \{3\} \\
A_1 &= \{1, 3\}, A_2 = \{2\} \\
A_1 &= \{1\}, A_2 = \{2, 3\} \\
A_1 &= \{1\}, A_2 = \{2\}, A_3 = \{3\}
\end{aligned}$$

Let $g(n)$ be the number of partitions of a set with n elements.
(a) Show that $g(n) = \sum_{k=0}^{n-1} \binom{n-1}{k} g(k)$ [define $g(0) = 1$].
(b) Show that $g(n) = e^{-1} \sum_{k=0}^{\infty} k^n/k!$
HINT: Show that the series satisfies the difference equation of part (a).

1.5 INDEPENDENCE

Consider the following experiment. A person is selected at random and his height is recorded. After this the last digit of the license number of the next car to pass is noted. If A is the event that the height is over 6 feet, and B is the event that the digit is ≥ 7, then, intuitively, A and B are "independent" in the sense that knowledge about the occurrence or nonoccurrence of one of the events should not influence the odds about the other. For example, say that $P(A) = .2$, $P(B) = .3$. In a long sequence of trials we would expect the following situation.

(Roughly) 20% of the time A occurs; of those cases in which A occurs:
30% B occurs
70% B does not occur

80% of the time A does not occur; of these cases:
30% B occurs
70% B does not occur

Thus, if B is independent of A, it appears that $P(A \cap B)$ should be $.2(.3) = .06 = P(A)P(B)$, and $P(A^c \cap B)$ should be $.8(.3) = .24 = P(A^c)P(B)$.

Conversely, if $P(A \cap B) = P(A)P(B) = .06$ and $P(A^c \cap B) = P(A^c)P(B) = .24$, then, if A occurs roughly 20% of the time and we look at only the cases in which A occurs, B must occur in roughly 30% of these cases in order to have $A \cap B$ occur 6% of the time. Similarly, if we look at the cases in which A does not occur (80%), then, since we are assuming that $A^c \cap B$ occurs 24% of the time, we must have B occurring in 30% of these cases. Thus the odds about B are not changed by specifying the occurrence or nonoccurrence of A.

It appears that we should say that event B is independent of A iff $P(A \cap B) = P(A)P(B)$ and $P(A^c \cap B) = P(A^c)P(B)$. However, the second condition is already implied by the first. If $P(A \cap B) = P(A)P(B)$,

$$P(A^c \cap B) = P(B - A) = P(B - (A \cap B)) = P(B) - P(A \cap B)$$

since $A \cap B$ is a subset of B; hence

$$P(A^c \cap B) = P(B) - P(A)P(B) = (1 - P(A))P(B) = P(A^c)P(B)$$

Thus B is independent of A; that is, knowledge of A does not influence the odds about B, iff $P(A)P(B) = P(A \cap B)$. But this condition is perfectly symmetrical, in other words, B is independent of A iff A is independent of B. Thus we are led to the following definition.

Definition. Two events A and B are *independent* iff $P(A \cap B) = P(A)P(B)$.

If we have three events A, B, C that are (intuitively) independent, knowledge of the occurrence or nonoccurrence of $A \cap B$, for example, should not change the odds about C; this leads as above to the requirement that $P(A \cap B \cap C) = P(A \cap B)P(C)$. But if A, B, and C are to be independent, we must expect that A and B are independent (as well as A and C, and B. and C), so we should have all of the following conditions satisfied.

$$P(A \cap B) = P(A)P(B), \qquad P(A \cap C) = P(A)P(C),$$
$$P(B \cap C) = P(B)P(C)$$

and

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

We are led to the following definition.

Definition. Let A_i, $i \in I$, where I is an arbitrary index set, possibly infinite, be an arbitrary collection of events [a fixed probability space $(\Omega, \mathscr{F}, P)$ is of course assumed].

The A_i are said to be *independent* iff for each finite set of distinct indices $i_1, \dots, i_k \in I$ we have

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_k})$$

REMARKS

1. If the A_i, $i \in I$, are independent, it follows that

$$P(B_{i_1} \cap \cdots \cap B_{i_k}) = P(B_{i_1}) \cdots P(B_{i_k})$$

for all (distinct) $i_1, \dots, i_k$, where each B_{i_r} may be either A_{i_r} or $A_{i_r}{}^c$. To put it simply, if the A_i are independent and we replace any event by its complement, we still have independence [see Problem 1; actually we have already done most of the work by showing that $P(A \cap B) = P(A)P(B)$ implies $P(A^c \cap B) = P(A^c)P(B)$].

2. The condition $P(A_1 \cap \cdots \cap A_n) = P(A_1) \cdots P(A_n)$ does not imply the analogous condition for any smaller family of events. For example, it is possible to have $P(A \cap B \cap C) = P(A)P(B)P(C)$, but $P(A \cap B) \neq P(A)P(B)$, $P(A \cap C) \neq P(A)P(C)$, $P(B \cap C) \neq P(B)P(C)$. In particular, A, B, and C are not independent.

Conversely it is possible to have, for example, $P(A \cap B) = P(A)P(B)$, $P(A \cap C) = P(A)P(C)$, $P(B \cap C) = P(B)P(C)$, but $P(A \cap B \cap C) \neq P(A)P(B)P(C)$. Thus A and B are independent, as are A and C, and also B and C, but A, B, and C are not independent.

▶ **Example 1.** Let two dice be tossed, and take $\Omega =$ all ordered pairs (i, j), $i, j = 1, 2, \dots, 6$, with each point assigned probability $1/36$. Let

$$A = \{\text{first die} = 1, 2, \text{ or } 3\}$$

$$B = \{\text{first die} = 3, 4, \text{ or } 5\}$$

$$C = \{\text{the sum of the two faces is } 9\}$$

(Thus $A \cap B = \{(3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6)\}$, $A \cap C = \{(3, 6)\}$, $B \cap C = \{(3, 6), (4, 5), (5, 4)\}$, $A \cap B \cap C = \{(3, 6)\}$.) Then

$$P(A \cap B) = \tfrac{1}{6} \neq P(A)P(B) = \tfrac{1}{2}(\tfrac{1}{2}) = \tfrac{1}{4}$$

$$P(A \cap C) = \tfrac{1}{36} \neq P(A)P(C) = \tfrac{1}{2}(\tfrac{4}{36}) = \tfrac{1}{18}$$

$$P(B \cap C) = \tfrac{1}{12} \neq P(B)P(C) = \tfrac{1}{2}(\tfrac{1}{9}) = \tfrac{1}{18}$$

But

$$P(A \cap B \cap C) = \tfrac{1}{36} = P(A)P(B)P(C)$$

Now in the same probability space let

$$A = \{\text{first die} = 1, 2, \text{ or } 3\}$$
$$B = \{\text{second die} = 4, 5, \text{ or } 6\}$$
$$C = \{\text{the sum of the two faces is } 7\}$$

(Thus $A \cap C = \{(1, 6), (2, 5), (3, 4)\} = A \cap B \cap C$, etc.) Then

$$P(A \cap B) = \tfrac{1}{4} = P(A)P(B) = \tfrac{1}{2}(\tfrac{1}{2})$$
$$P(A \cap C) = \tfrac{1}{12} = P(A)P(C) = \tfrac{1}{2}(\tfrac{1}{6})$$
$$P(B \cap C) = \tfrac{1}{12} = P(B)P(C) = \tfrac{1}{2}(\tfrac{1}{6})$$

But

$$P(A \cap B \cap C) = \tfrac{1}{12} \neq P(A)P(B)P(C) = \tfrac{1}{24} \blacktriangleleft$$

We illustrate the idea of independence by considering some problems related to the classical coin-tossing experiment.

A sequence of *n Bernoulli trials* is a sequence of n independent observations, each of which may result in exactly one of two possible situations, called "success" or "failure." At each observation the probability of success is p, and the probability of failure is $q = 1 - p$.

Special Cases

(a) Toss a coin independently n times, with success = heads, failure = tails.

(b) Examine components produced on an assembly line; success = acceptable, failure = defective.

(c) Transmit binary digits through a communication channel; success = digit recieved correctly, failure = digit received incorrectly.

We take Ω = all 2^n ordered sequences of length n, with components 0 (failure) and 1 (success). To assign probabilities in accordance with the physical description given above, we reason as follows.

Consider the sample point $\omega = 11 \cdots 10 \cdots 0$ (k 1's followed by $n - k$ 0's). Let A_i = {success on trial i} = the set of all sequences with a 1 in the ith coordinate. Because of the independence of the trials we must assign

$$\begin{aligned} P\{\omega\} &= P(A_1 \cap A_2 \cap \cdots \cap A_k \cap A_{k+1}^c \cap \cdots \cap A_n{}^c) \\ &= P(A_1)P(A_2) \cdots P(A_k)P(A_{k+1}^c) \cdots P(A_n{}^c) = p^k q^{n-k} \end{aligned}$$

Similarly, *any point with k 1's and n − k 0's is assigned probability* $p^k q^{n-k}$.

The number of such points is the number of ways of selecting k distinct positions for the 1's to occur (or selecting $n - k$ distinct positions for the 0's); that is, $\binom{n}{k}$. The sum of the probabilities assigned to all the points is

$$\sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} = (p + q)^n = 1$$

by the binomial theorem. Thus we have a legitimate assignment. Furthermore, the probability of obtaining exactly k successes is

$$p(k) = \binom{n}{k} p^k q^{n-k} \qquad k = 0, 1, \dots, n \tag{1.5.1}$$

$p(k)$, $k = 0, 1, \dots, n$, is called the *binomial* probability function.

▶ **Example 2.** Six balls are tossed independently into three boxes A, B, C. For each ball the probability of going into a specific box is 1/3. Find the probability that box A will contain (a) exactly four balls, (b) at least two balls, (c) at least five balls.

Here we have six Bernoulli trials, with success corresponding to a ball in box A, failure to a ball in box B or C. Thus $n = 6$, $p = 1/3$, $q = 2/3$, and so the required probabilities are

(a) $p(4) = \binom{6}{4}(\frac{1}{3})^4(\frac{2}{3})^2$

(b) $1 - p(0) - p(1) = 1 - (\frac{2}{3})^6 - \binom{6}{1}(\frac{1}{3})(\frac{2}{3})^5$

(c) $p(5) + p(6) = \binom{6}{5}(\frac{1}{3})^5(\frac{2}{3}) + (\frac{1}{3})^6$ ◀

We now consider *generalized Bernoullli trials.* Here we have a sequence of independent trials, and on each trial the result is exactly one of the k possibilities $b_1, \dots, b_k$. On a given trial let b_i occur with probability p_i, $i = 1, 2, \dots, k$ ($p_i \geq 0$, $\sum_{i=1}^{k} p_i = 1$).

We take $\Omega =$ all k^n ordered sequences of length n with components $b_1, \dots, b_k$; for example, if $\omega = (b_1 b_3 b_2 b_2 \cdots)$ then b_1 occurs on trial 1, b_3 on trial 2, b_2 on trials 3 and 4, and so on. As in the previous situation, assign to the point

$$\omega = (\underbrace{b_1 b_1 \cdots b_1}_{n_1} \underbrace{b_2 \cdots b_2}_{n_2} \cdots \underbrace{b_k \cdots b_k}_{n_k})$$

the probability $p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k}$. This is the probability assigned to any sequence having n_i occurrences of b_i, $i = 1, 2, \dots, k$. To find the number of such sequences, first select n_1 positions out of n for the b_1's to occur, then n_2 positions out of the remaining $n - n_1$ for the b_2's, n_3 out of $n - n_1 - n_2$ for the b_3's, and so on. Thus the number of sequences having exactly n_1

occurrences of $b_1, \ldots, n_k$ occurrences of b_k is

$$\binom{n}{n_1}\binom{n-n_1}{n_2}\binom{n-n_1-n_2}{n_3}\cdots\binom{n-n_1-\cdots-n_{k-2}}{n_{k-1}}\binom{n_k}{n_k} = \frac{n!}{n_1!\,n_2!\cdots n_k!}$$

The total probability assigned to all points is

$$\sum_{\substack{n_1,\ldots,n_k \text{ nonneg} \\ \text{integers, with } \sum_{i=1}^k n_i = n}} \frac{n!}{n_1!n_2!\cdots n_k!} p_1^{n_1} p_2^{n_2}\cdots p_k^{n_k} = (p_1+\cdots+p_k)^n = 1 \tag{1.5.2}$$

To see this, notice that $(p_1+\cdots+p_k)^n = (p_1+\cdots+p_k)(p_1+\cdots+p_k)\cdots(p_1+\cdots+p_k)$, n times. A typical term in the expansion is $p_1^{n_1}\cdots p_k^{n_k}$; the number of times this term appears is

$$\frac{n!}{n_1!\,n_2!\cdots n_k!}$$

since we may count the appearances by selecting p_1 from n_1 of the n factors, selecting p_2 from n_2 of the remaining $n - n_1$ factors, and so forth. Thus we have a legitimate assignment of probabilities.

The probability that b_1 will occur n_1 times, b_2 will occur n_2 times, $\ldots$, and b_k will occur n_k times is

$$p(n_1, \ldots, n_k) = \frac{n!}{n_1!\cdots n_k!} p_1^{n_1}\cdots p_k^{n_k} \tag{1.5.3}$$

$p(n_1, \ldots, n_k)$, $n_1, \ldots, n_k$ = nonnegative integers whose sum is n, is called the *multinomial* probability function.

Note that when $k = 2$, generalized Bernoulli trials reduce to ordinary Bernoulli trials [let b_1 = "success," b_2 = "failure", $p_1 = p, p_2 = q = 1 - p$, $n_1 = k$, $n_2 = n - k$; then $(n!/n_1!\,n_2!)p_1^{n_1}p_2^{n_2} = \binom{n}{k}p^k q^{n-k}$ = probability of k successes in n trials].

▶ **Example 3.** Throw four unbiased dice independently. Find the probability of exactly two 1's and one 2.

Let

b_1 = "1 occurs" (on a *given* trial) $\quad p_1 = \frac{1}{6} \quad n_1 = 2$
b_2 = "2 occurs" $\quad p_2 = \frac{1}{6} \quad n_2 = 1,$
b_3 = "3, 4, 5, or 6 occurs" $\quad p_3 = \frac{2}{3} \quad n_3 = 1$

The probability is $(4!/2!\,1!\,1!)\,(1/6)^2(1/6)^1(2/3)^1 = 1/27$. ◀

▶ **Example 4.** If 10 balls are tossed independently into five boxes, with a given ball equally likely to fall into each box, find the probability that all boxes will have the same number of balls.

Let $b_i =$ "ball goes into box i", $p_i = 1/5$, $n_i = 2$, $i = 1, 2, 3, 4, 5$, $n = 10$. The probability $= (10!/2^5)(1/5)^{10}$. ◀

REMARK. If r balls are tossed independently into n boxes, we have seen that the event {ball 1 into box i_1, ball 2 into box $i_2, \ldots,$ ball r into box i_r} must have probability $p_{i_1}p_{i_2}\cdots p_{i_r}$, where p_i is the probability that a specific ball will fall into box i. In particular, if all $p_i = 1/n$ (as is assumed in Examples 2 and 4 above), the probability of the event is $1/n^r$. In other words, all ordered samples of size r (out of n symbols), with replacement, have the same probability. This justifies the assertion we made in Example 3 of Section 1.4.

We emphasize that the independence of the tosses is an *assumption*, not a theorem. For example, if two balls are tossed into two boxes and a given box can contain at most one ball, then the events $A_1 =$ {ball 1 goes into box 1} and $A_2 =$ {ball 2 goes into box 1} are not independent, since $P(A_1 \cap A_2) = 0$, $P(A_1) = P(A_2) = 1/2$.

▶ **Example 5.** An urn contains equal numbers of black, white, red, and green balls. Four balls are drawn independently, with replacement. Find the probability $p(k)$ that exactly k colors will appear in the sample, $k = 1, 2, 3, 4$.

This is a multinomial problem with $n = 4$ and $b_1 = B =$ black, $b_2 = W =$ white, $b_3 = R =$ red, $b_4 = G =$ green.

$k = 4$: The probability that all four colors will appear is given by the multinomial formula with all $n_i = 1$; that is,

$$\frac{4!}{1!\,1!\,1!\,1!}\left(\frac{1}{4}\right)^4 = 6/64 = p(4)$$

$k = 3$: The probability of obtaining two black, one white, and one red ball is given by the multinomial formula with $n_1 = 2$, $n_2 = n_3 = 1$, $n_4 = 0$; that is,

$$\frac{4!}{2!\,1!\,1!\,0!}\left(\frac{1}{4}\right)^4 = \frac{3}{64}$$

To find the total probability of obtaining exactly three colors, multiply by the number of ways of selecting three colors out of four $[\binom{4}{3} = 4]$ and the number of ways of selecting one of three colors to be repeated (3). Thus

$$p(3) = 36/64.$$

$k = 2$: The probability of obtaining two black and two white balls is

$$\frac{4!}{2!\,2!\,0!\,0!}\left(\frac{1}{4}\right)^4 = \frac{3}{128}$$

Thus the probability of obtaining two balls of one color and two of another is 3/128 times the number of ways of selecting two colors out of $4[\binom{4}{2} = 6]$, or 9/64. Similarly, the probability of obtaining three of one color and one of another is

$$\frac{4!}{3!\,1!\,0!\,0!}\left(\frac{1}{4}\right)^4(4)(3) = \frac{12}{64}$$

Notice that the extra factor is $(4)(3) = 12$, not $\binom{4}{2} = 6$, since three blacks and one white constitute a different selection from three whites and one black. Thus

$$p(2) = 9/64 + 12/64 = 21/64.$$

$k = 1$: The probability that all balls will be of the same color is

$$p(1) = \frac{4!}{4!\,0!\,0!\,0!}\left(\frac{1}{4}\right)^4(4) = \frac{1}{64}$$

REMARK. The sample space of this problem consists of all *ordered* samples of size 4, with replacement, from the symbols B, W, R, G, with all samples assigned the same probability. The reader should resist the temptation to assign equal probability to all unordered samples of size 4, with replacement. This would imply, for example, that $\{WWWW\}$ and $\{WWWB, WWBW, WBWW, BWWW\}$ = {three whites and one black} have the same probability, and this is inconsistent with the assumption of independence. ◄

PROBLEMS

1. Show that the events A_i, $i \in I$, are independent iff $P(B_{i_1} \cap \cdots \cap B_{i_k}) = P(B_{i_1}) \cdots P(B_{i_k})$ for all (distinct) $i_1, \ldots, i_k$, where each B_{i_r} may be either A_{i_r} or $A_{i_r}{}^c$.
2. Let $p(k)$, $k = 0, 1, \ldots, n$, be the binomial probability function.
   (a) If $(n + 1)p$ is not an integer, show that $p(k)$ is strictly increasing up to $k = [(n + 1)p]$ = the largest integer $\leq (n + 1)p$, and attains a maximum at $[(n + 1)p]$. $p(k)$ is strictly decreasing for all larger values of k.
   (b) If $(n + 1)p$ is an integer, show that $p(k)$ is strictly increasing up to $k = (n + 1)p - 1$ and has a double maximum at $k = (n + 1)p - 1$ and $k = (n + 1)p$; $p(k)$ is strictly decreasing for larger values of k.

3. A single card is drawn from an ordinary deck. Give examples of events A and B associated with this experiment that are
(a) Mutually exclusive (disjoint) but not independent
(b) Independent but not mutually exclusive
(c) Independent and mutually exclusive
(d) Neither independent nor mutually exclusive

4. Of the 100 people in a certain village, 50 always tell the truth, 30 always lie, and 20 always refuse to answer. A sample of size 30 is taken with replacement.
(a) Find the probability that the sample will contain 10 people of each category.
(b) Find the probability that there will be exactly 12 liars.

5. Six unbiased dice are tossed independently. Find the probability that the number of 1's minus the number of 2's will be 3.

6. How many terms are there in the multinomial expansion (1.5.2)?

7. An urn contains t_1 balls of color C_1, t_2 of color $C_2, \ldots, t_k$ of color C_k.
(a) If n balls are drawn without replacement, show that the probability of obtaining exactly n_1 of color C_1, n_2 of color $C_2, \ldots, n_k$ of color C_k is

$$\frac{\binom{t_1}{n_1}\binom{t_2}{n_2}\cdots\binom{t_k}{n_k}}{\binom{t}{n}}$$

where $t = t_1 + t_2 + \cdots + t_k$ is the total number of balls in the urn and $\binom{t_i}{n_i}$ is defined to be 0 if $n_i > t_i$. (Notice the pattern: $t_1 + t_2 + \cdots + t_k = t$, $n_1 + n_2 + \cdots + n_k = n$.) The above expression, regarded as a function of $n_1, \ldots, n_k$, is called the *hypergeometric* probability function.
(b) What is the probability of the event of part (a) if the balls are drawn independently, with replacement?

8. (a) If an event A is independent of itself, that is, if A and A are independent, show that $P(A) = 0$ or 1.
(b) If $P(A) = 0$ or 1, show that A and B are independent for any event B, in particular, that A and A are independent.

1.6 CONDITIONAL PROBABILITY

If A and B are independent events, the occurrence or nonoccurrence of A does not influence the odds concerning the occurrence of B. If A and B are not independent, it would be desirable to have some way of measuring exactly how much the occurrence of one of the events changes the odds about the other.

In a long sequence of independent repetitions of the experiment, $P(A)$ measures the fraction of the trials on which A occurs. If we look only at the trials on which A occurs (say there are n_A of these) and record those trials on which B occurs also (there are n_{AB} of these, where n_{AB} is the number of trials on which both A and B occur), the ratio n_{AB}/n_A is a measure of $P(B \mid A)$, the "conditional probability of B given A," that is, the fraction of the time that B occurs, looking only at trials producing an occurrence of A. Comparing $P(B \mid A)$ with $P(B)$ will indicate the difference between the odds about B when A is known to have occurred, and the odds about B before any information about A is revealed.

The above discussion suggests that we define the *conditional probability of B given A* as

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} \tag{1.6.1}$$

This makes sense if $P(A) > 0$.

▶ **Example 1.** Throw two unbiased dice independently. Let $A = \{\text{sum of the faces} = 8\}$, $B = \{\text{faces are equal}\}$. Then

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{P\{4-4\}}{P\{4-4, 5-3, 3-5, 6-2, 2-6\}} = \frac{1/36}{5/36} = \frac{1}{5}$$

(see Figure 1.6.1).

There is a point here that may be puzzling. In counting the outcomes favorable to A, we note that there are two ways of making an 8 using a 5 and a 3, but only one way using a 4 and a 4. The probability space consists of all 36 *ordered* pairs (i, j), $i, j = 1, 2, 3, 4, 5, 6$, each assigned probability$1/36$. The ordered pair (4, 4) is the same as the ordered pair (4, 4) (this is rather difficult to dispute), while (5, 3) is different from (3, 5). Alternatively, think of the first die to be thrown as red and the second as green. A 5 on the red die and a 3 on the green is a different outcome from a 3 on the red and a 5 on the green. However, using 4's we can make an 8 in only one way, a 4 on the red followed by a 4 on the green. ◀

The extension of the definition of conditional probability to events with probability zero will be considered in great detail later on. For now, we are

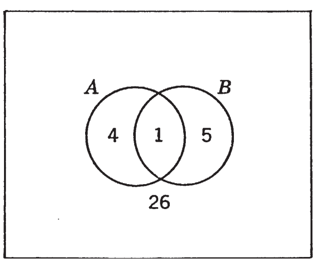


FIGURE 1.6.1 Example on Conditional Probability. Numbers Indicate Favorable Outcomes.

content to note some consequences of the above definition [whenever an expression such as $P(B \mid A)$ is written, it is assumed that $P(A) > 0$].

If A and B are independent, then $P(A \cap B) = P(A)P(B)$, so that $P(B \mid A) = P(B)$ and $P(A \mid B)$ $(= P(A \cap B)/P(B)) = P(A)$, in accordance with the intuitive notion that the occurrence of one of the events does not change the odds about the other.

The formula $P(A \cap B) = P(A)P(B \mid A)$ may be extended to more than two events.

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

Hence

$$P(A \cap B \cap C) = P(A)P(B \mid A)P(C \mid A \cap B) \tag{1.6.2}$$

Similarly

$$P(A \cap B \cap C \cap D) = P(A)P(B \mid A)P(C \mid A \cap B)P(D \mid A \cap B \cap C) \tag{1.6.3}$$

and so on.

▶ **Example 2.** Three cards are drawn without replacement from an ordinary deck. Find the probability of not obtaining a heart.

Let $A_i = \{\text{card } i \text{ is not a heart}\}$. Then we are looking for

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1 \cap A_2) = \tfrac{39}{52}\tfrac{38}{51}\tfrac{37}{50}$$

[For example, to find $P(A_2 \mid A_1)$, we restrict ourselves to the outcomes favorable to A_1. If the first card is not a heart, 51 cards remain in the deck, including 13 hearts, so that the probability of not getting a heart on the second trial is 38/51.]

Notice that the above probability can be written $\binom{39}{3}/\binom{52}{3}$, which could have been derived by direct combinatorial reasoning. Furthermore, if the cards were drawn independently, with replacement, the probability would be quite different, $(3/4)^3 = 27/64$. ◀

We now prove one of the most useful theorems of the subject.

Theorem of Total Probability. *Let $B_1, B_2, \ldots$ be a finite or countably infinite family of mutually exclusive and exhaustive events (i.e., the B_i are disjoint and their union is Ω). If A is any event, then*

$$P(A) = \sum_i P(A \cap B_i) \tag{1.6.4}$$

Thus $P(A)$ is computed by finding a list of mutually exclusive, exhaustive ways in which A can happen, and then adding the individual probabilities.

Also

$$P(A) = \sum_i P(B_i)P(A \mid B_i) \tag{1.6.5}$$

where the sum is taken over those i for which $P(B_i) > 0$. Thus $P(A)$ is a weighted average of the conditional probabilities $P(A \mid B_i)$.

PROOF.

$$P(A) = P(A \cap \Omega) = P\left(A \cap \left(\bigcup_i B_i\right)\right) = P\left(\bigcup_i (A \cap B_i)\right)$$
$$= \sum_i P(A \cap B_i) = \sum_i P(B_i)P(A \mid B_i)$$

Notice that under the above assumptions we have

$$P(B_k \mid A) = \frac{P(A \cap B_k)}{P(A)} = \frac{P(B_k)P(A \mid B_k)}{\sum_i P(B_i)P(A \mid B_i)} \tag{1.6.6}$$

This formula is sometimes referred to as *Bayes' theorem*; $P(B_k \mid A)$ is sometimes called an *a posteriori probability*. The reason for this terminology may be seen in the example below.

▶ **Example 3.** Two coins are available, one unbiased and the other two-headed. Choose a coin at random and toss it once; assume that the unbiased coin is chosen with probability 3/4. Given that the result is heads, find the probability that the two-headed coin was chosen.

The "tree diagram" shown in Figure 1.6.2 represents the experiment.

We may take Ω to consist of the four possible paths through the tree, with each path assigned a probability equal to the product of the probabilities assigned to each branch. Notice that we are given the probabilities of the

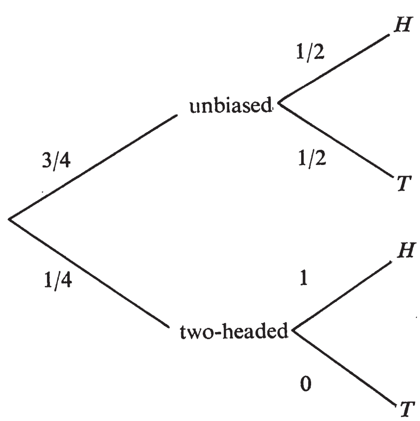


FIGURE 1.6.2 Tree Diagram.

events $B_1 = \{$unbiased coin chosen$\}$ and $B_2 = \{$two-headed coin chosen$\}$, as well as the conditional probabilities $P(A \mid B_i)$, where $A = \{$coin comes up heads$\}$. This is sufficient to determine the probabilities of all events.

Now we can compute $P(B_2 \mid A)$ using Bayes' theorem; this is facilitated if, instead of trying to identify the individual terms in (1.6.6), we simply look at the tree and write

$$\begin{aligned} P(B_2 \mid A) &= \frac{P(B_2 \cap A)}{P(A)} \\ &= \frac{P\{\text{two-headed coin chosen and coin comes up heads}\}}{P\{\text{coin comes up heads}\}} \\ &= \frac{(1/4)(1)}{(3/4)(1/2) + (1/4)(1)} = \frac{2}{5} \blacktriangleleft \end{aligned}$$

There are many situations in which an experiment consists of a sequence of steps, and the conditional probabilities of events happening at step $n + 1$, given outcomes at step n, are specified. In such cases a description by means of a tree diagram may be very convenient (see Problems).

▶ **Example 4.** A loaded die is tossed once; if N is the result of the toss, then $P\{N = i\} = p_i$, $i = 1, 2, 3, 4, 5, 6$. If $N = i$, an unbiased coin is tossed independently i times. Find the conditional probability that N will be odd, given that at least one head is obtained (see Figure 1.6.3).

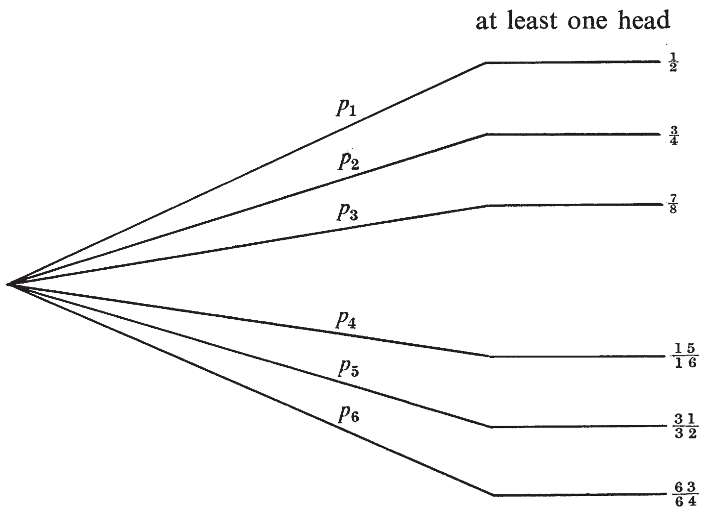


FIGURE 1.6.3

Let $A = \{\text{at least one head obtained}\}$, $B = \{N \text{ odd}\}$. Then $P(B \mid A) = P(A \cap B)/P(A)$. Now

$$P(A \cap B) = \sum_{i=1,3,5} P\{N = i \text{ and at least one head obtained}\}$$
$$= \tfrac{1}{2}p_1 + \tfrac{7}{8}p_3 + \tfrac{31}{32}p_5$$

since when an unbiased coin is tossed independently i times, the probability of at least one head is $1 - (1/2)^i$. Similarly,

$$P(A) = \sum_{i=1}^{6} P\{N = i \text{ and at least one head obtained}\}$$
$$= \sum_{i=1}^{6} p_i(1 - 2^{-i})$$

Thus

$$P(B \mid A) = \frac{\tfrac{1}{2}p_1 + \tfrac{7}{8}p_3 + \tfrac{31}{32}p_5}{\tfrac{1}{2}p_1 + \tfrac{3}{4}p_2 + \tfrac{7}{8}p_3 + \tfrac{15}{16}p_4 + \tfrac{31}{32}p_5 + \tfrac{63}{64}p_6} \quad ◀$$

PROBLEMS

1. In 10 Bernoulli trials find the conditional probability that all successes will occur consecutively (i.e., no two successes will be separated by one or more failures), given that the number of successes is between four and six.

2. If X is the number of successes in n Bernoulli trials, find the probability that $X \geq 3$ given that $X \geq 1$.

3. An unbiased die is tossed once. If the face is odd, an unbiased coin is tossed repeatedly; if the face is even, a biased coin with probability of heads $p \neq 1/2$ is tossed repeatedly. (Successive tosses of the coin are independent in each case.) If the first n throws result in heads, what is the probability that the unbiased coin is being used?

4. A positive integer I is selected, with $P\{I = n\} = (1/2)^n$, $n = 1, 2, \ldots$. If I takes the value n, a coin with probability of heads e^{-n} is tossed once. Find the probability that the resulting toss will be a head.

5. A bridge player and his partner are known to have six spades between them. Find the probability that the spades will be split

(a) 3-3
(b) 4-2 or 2-4
(c) 5-1 or 1-5
(d) 6-0 or 0-6.

6. An urn contains 30 white and 15 black balls. If 10 balls are drawn with (respectively without) replacement, find the probability that the first two balls will be white, given that the sample contains exactly six white balls.

7. Let C_1 be an unbiased coin, and C_2 a biased coin with probability of heads 3/4. At time $t = 0$, C_1 is tossed. If the result is heads, then C_1 is tossed at time $t = 1$; if the result is tails, C_2 is tossed at $t = 1$. The process is repeated at $t = 2, 3, \ldots$. In general, if heads appears at $t = n$, then C_1 is tossed at $t = n + 1$; if tails appears at $t = n$, then C_2 is tossed at $t = n + 1$.

Find $y_n =$ the probability that the toss at $t = n$ will be a head (set up a difference equation).

8. In the switching network of Figure P.1.6.8, the switches operate independently.

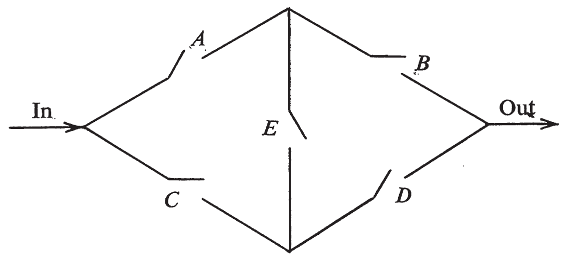


FIGURE P.1.6.8

Each switch closes with probability p, and remains open with probability $1 - p$.

(a) Find the probability that a signal at the input will be received at the output.

(b) Find the conditional probability that switch E is open given that a signal is received.

9. In a certain village 20% of the population has disease D. A test is administered which has the property that if a person has D, the test will be positive 90% of the time, and if he does not have D, the test will still be positive 30% of the time. All those whose test is positive are given a drug which invariably cures the disease, but produces a characteristic rash 25% of the time. Given that a person picked at random has the rash, what is the probability that he actually had D to begin with?

1.7 SOME FALLACIES IN COMBINATORIAL PROBLEMS

In this section we illustrate some common traps occurring in combinatorial problems. In the first three examples there will be a multiple count.

▶ **Example 1.** Three cards are selected from an ordinary deck, without replacement. Find the probability of not obtaining a heart.

PROPOSED SOLUTION. The total number of selections is $\binom{52}{3}$. To find the number of favorable outcomes, notice that the first card cannot be a heart; thus we have 39 choices at step 1. Having removed one card, there are 38 nonhearts left at step 2 (and then 37 at step 3). The desired probability is $(39)(38)(37)/\binom{52}{3}$.

FALLACY. In computing the number of favorable outcomes, a particular selection might be: 9 of diamonds, 8 of clubs, 7 of diamonds. Another selection is: 8 of clubs, 9 of diamonds, 7 of diamonds. In fact the $3! = 6$ possible orderings of these three cards are counted separately in the numerator (but not in the denominator). Thus the proposed answer is too high by a factor of $3!$; the actual probability is $(39)(38)(37)/3!\binom{52}{3} = \binom{39}{3}/\binom{52}{3}$ (see example 2, Section 1.6). ◀

▶ **Example 2.** Find the probability that a five-card poker hand will result in three of a kind (three cards of the same face value x, plus two cards of face values y and z, with x, y, and z distinct).

PROPOSED SOLUTION. Pick the face value to appear three times (13 possibilities). Pick three suits out of four for the "three of a kind" ($\binom{4}{3}$ choices). Now one face value is excluded, so that 48 cards are left in the deck. Pick one of them as the fourth card; the fifth card can be chosen in 44 ways, since the fourth card excludes another face value. Thus the desired probability is $(13)\binom{4}{3}(48)(44)/\binom{52}{5}$.

FALLACY. Say the first three cards are aces. The fourth and fifth cards might be the jack of clubs and the 6 of diamonds, or equally well the 6 of diamonds and the jack of clubs. These possibilities are counted separately in the numerator but not in the denominator, so that the proposed answer is too high by a factor of 2. The actual probability is $13\binom{4}{3}(48)(44)/2\binom{52}{5} = 13\binom{4}{3}\binom{12}{2}16/\binom{52}{5}$ [see Problem 2, Section 1.4; the factor $\binom{12}{2}16$ corresponds to the selection of two distinct face values out of the remaining 12, then one card from each of these face values].

REMARK. A more complicated approach to this problem is as follows. Pick the face value x to appear three times, then pick three suits out of four, as before. Forty-nine cards remain in the deck, and the total number of ways of selecting two remaining cards is $\binom{49}{2}$. However, if the two face values are the same, we obtain a full house; there are $12\binom{4}{2}$ selections in which this happens (select one face value out of 12, then two suits out of four). Also, if one of the two cards has face value x, we obtain four of a kind; since there is only one remaining card with face value x and 48 cards remain after this one is chosen, there are 48 possibilities. Thus the probability of obtaining three of a kind is

$$\frac{13\binom{4}{3}\left[\binom{49}{2} - 12\binom{4}{2} - 48\right]}{\binom{52}{5}}$$

(This agrees with the previous answer.) ◀

▶ **Example 3.** Ten cards are drawn without replacement from an ordinary deck. Find the probability that at least nine will be of the same suit.

PROPOSED SOLUTION. Pick the suit in any one of four ways, then choose nine of 13 face values. Forty-three cards now remain in the deck, so that the desired probability is $4\binom{13}{9}43/\binom{52}{10}$.

FALLACY. Consider two possible selections.

1. Spades are chosen, then face values A K Q J 10 9 8 7 6. The last card is the 5 of spades.

2. Spades are chosen, then face values A K Q J 10 9 8 7 5. The last card is the 6 of spades (see Figure 1.7.1). Both selections yield the same 10 cards, but are counted separately in the computation. To find the number of duplications, notice that we can select 10 cards out of 13 to be involved in the duplication; each choice of one card (out of 10) for the last card yields a distinct path in Figure 1.7.1. Of the 10 possible paths corresponding to a given selection of cards, nine are redundant. Thus the actual probability is

$$\frac{4\left[\binom{13}{9}43 - \binom{13}{10}9\right]}{\binom{52}{10}}$$

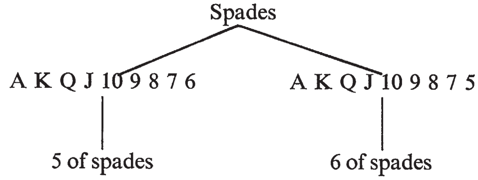


FIGURE 1.7.1 Multiple Count.

Now

$$\begin{aligned}\binom{13}{9}43 - \binom{13}{10}9 &= \binom{13}{9}39 + \binom{13}{9}4 - \binom{13}{10}9\\ &= \binom{13}{9}39 + \binom{13}{10}10 - \binom{13}{10}9\\ &= \binom{13}{9}39 + \binom{13}{10}\end{aligned}$$

so that the probability is

$$\frac{4\left[\binom{13}{9}39 + \binom{13}{10}\right]}{\binom{52}{10}}$$

as obtained in a straightforward manner in Problem 4, Section 1.4. ◄

▶ **Example 4.** An urn contains 10 balls $b_1, \ldots, b_{10}$. Five balls are drawn without replacement. Find the probability that b_8 and b_9 will be included in the sample.

PROPOSED SOLUTION. We are drawing half the balls, so that the probability that a particular ball will be included is 1/2. Thus the probability of including both b_8 and b_9 is $(1/2)(1/2) = 1/4$.

FALLACY. Let $A = \{b_8 \text{ is included}\}$, $B = \{b_9 \text{ is included}\}$. The difficulty is simply that A and B are not independent. For $P(A \cap B) = \binom{8}{3}/\binom{10}{5} = 2/9$ (after b_8 and b_9 are chosen, three balls are to be selected from the remaining eight). Also $P(A) = P(B) = \binom{9}{4}/\binom{10}{5} = 1/2$, so that $P(A \cap B) \neq P(A)P(B)$. ◄

▶ **Example 5.** Two cards are drawn independently, with replacement, from an ordinary deck; at each selection all 52 cards are equally likely. Find the probability that the king of spades and the king of hearts will be chosen (in some order).

PROPOSED SOLUTION. The number of unordered samples of size 2 out of 52, with replacement, is $\binom{52+2-1}{2} = \binom{53}{2}$ [see (1.4.4)]. The kings of spades and hearts constitute one such sample, so that the desired probability is $1/\binom{53}{2}$.

FALLACY. It is not legitimate to assign equal probability to all unordered samples with replacement. If we do this we are saying, for example, that the outcomes "ace of spades, ace of spades" and "king of spades, king of hearts" have the same probability. However, this cannot be the case if independent sampling is assumed. For the probability that the ace of spades is chosen twice is $(1/52)^2$, while the probability that the spade and heart kings will be chosen (in some order) is $P\{$first card is the king of spades, second card is the king of hearts$\}$ + $P\{$first card is the king of hearts, second card is the king of spades$\} = 2(1/52)^2$, which is the desired probability.

The main point is that we must use *ordered* samples with replacement in order to capture the idea of independence. ◀

1.8 APPENDIX: STIRLING'S FORMULA

An estimate of $n!$ that is of importance both in numerical calculations and theoretical analysis is *Stirling's formula*

$$n! \sim n^n e^{-n} \sqrt{2\pi n}$$

in the sense that

$$\lim_{n\to\infty} \frac{n!}{(n^n e^{-n}\sqrt{2\pi n})} = 1$$

PROOF. Define $(2n)!!$ (read $2n$ semifactorial) as $2n(2n-2)(2n-4)\cdots 6(4)(2)$, and $(2n+1)!!$ as $(2n+1)(2n-1)\cdots(5)(3)(1)$. We first show that

$$\text{(a)}\qquad \frac{(2n)!!}{(2n+1)!!} < \frac{\pi}{2}\frac{(2n-1)!!}{(2n)!!} < \frac{(2n-2)!!}{(2n-1)!!}$$

Let $I_k = \int_0^{\pi/2} (\cos x)^k\, dx$, $k = 0, 1, 2, \ldots$. Then $I_0 = \pi/2$, $I_1 = 1$. Integrating by parts, we obtain $I_k = \int_0^{\pi/2} (\cos x)^{k-1}\, d(\sin x) = \int_0^{\pi/2} (k-1)(\cos x)^{k-2} \sin^2 x\, dx$. Since $\sin^2 x = 1 - \cos^2 x$, we have $I_k = (k-1)I_{k-2} - (k-1)I_k$ or $I_k = [(k-1)/k]I_{k-2}$. By iteration, we obtain $I_{2n} = (\pi/2)\,[(2n-1)!!/(2n)!!]$ and $I_{2n+1} = [(2n)!!/(2n+1)!!]$. Since $(\cos x)^k$ decreases with k, so does I_k, and hence $I_{2n+1} < I_{2n} < I_{2n-1}$, and (a) is proved.

(b) Let $Q_n = \binom{2n}{n}/2^{2n}$. Then

$$\lim_{n\to\infty} Q_n \sqrt{n\pi} = 1$$

To prove this, write

$$Q_n = \frac{(2n)!}{n!\,n!\,2^{2n}} = \frac{(2n)!}{(2^n n!)^2}$$

$$= \frac{(2n)!}{((2n)(2n-2)\cdots(4)(2))^2} = \frac{(2n-1)!!}{(2n)!!}$$

Thus, by (a),

$$\frac{(2n)!!}{(2n+1)!!} < \frac{\pi}{2} Q_n < \frac{(2n-2)!!}{(2n-1)!!}$$

Multiply this inequality by

$$\frac{(2n-1)!!}{(2n-2)!!} = \frac{(2n-1)!!}{(2n)!!} \frac{(2n)!!}{(2n-2)!!} = Q_n(2n)$$

to obtain

$$\frac{2n}{2n+1} < n\pi Q_n^{\,2} < 1$$

If we let $n \to \infty$, we obtain $n\pi Q_n^{\,2} \to 1$, proving (b).

(c) *Proof of Stirling's formula.* Let $c_n = n!/n^n e^{-n}\sqrt{2\pi n}$. We must show that $c_n \to 1$ as $n \to \infty$. Consider $(n+1)!/n! = n+1$. We have

$$\frac{(n+1)!}{n!} = \frac{c_{n+1}(n+1)^{n+1}e^{-(n+1)}\sqrt{2\pi(n+1)}}{c_n n^n e^{-n}\sqrt{2\pi n}}$$

$$= \left(\frac{c_{n+1}}{c_n}\right) e^{-1} \left(\frac{n+1}{n}\right)^n \frac{(n+1)^{3/2}}{\sqrt{n}}$$

Thus

$$\frac{c_{n+1}}{c_n} = (n+1)(e)\left(\frac{n}{n+1}\right)^n \frac{\sqrt{n}}{(n+1)^{3/2}} = (e)\left(1+\frac{1}{n}\right)^{-(n+1/2)}$$

Now $(1 + 1/n)^{n+1/2} > e$ for n sufficiently large (take logarithms and expand in a power series); hence $c_{n+1}/c_n < 1$ for large enough n. Since every monotone bounded sequence converges, $c_n \to$ a limit c. We must show $c = 1$. By (b),

$$\lim_{n\to\infty} \binom{2n}{n} \sqrt{n\pi}\, 2^{-2n} = 1$$

But

$$\binom{2n}{n}\sqrt{n\pi}\, 2^{-2n} = \frac{(2n)!}{n!\,n!}\frac{\sqrt{n\pi}}{2^{2n}} = \frac{c_{2n}(2n/e)^{2n}\sqrt{2\pi(2n)}}{(c_n(n/e)^n\sqrt{2\pi n})^2}\frac{\sqrt{n\pi}}{2^{2n}} = \frac{c_{2n}}{c_n^{\,2}}$$

Therefore $c_{2n}/c_n^{\,2} \to 1$. However, $c_{2n} \to c$ and $c_n^{\,2} \to c^2$, and consequently $c/c^2 = 1$, so that $c = 1$. The theorem is proved.

Remark. The last step requires that c be >0. To see this, write

$$c_{n+1} = \frac{c_1}{c_0}\frac{c_2}{c_1}\cdots\frac{c_{n+1}}{c_n}$$

where c_0 is defined as 1. To show that $c_n \to a$ nonzero limit, it suffices to show that the limit of $\ln c_{n+1}$ is finite, and for this it is sufficient to show that $\sum_n \ln (c_{n+1}/c_n)$ converges to a finite limit. Now

$$\ln \frac{c_{n+1}}{c_n} = \ln \left[e\left(1 + \frac{1}{n}\right)^{-(n+1/2)} \right] = 1 - (n + \tfrac{1}{2}) \ln \left(1 + \frac{1}{n}\right)$$

$$= 1 - (n + \tfrac{1}{2})\left(\frac{1}{n} - \frac{1}{2n^2} + \frac{\theta(n)}{n^3}\right)$$

where $\theta(n)$ is bounded by a constant independent of n. This is the order of $1/n^2$; hence $\sum_n \ln (c_{n+1}/c_n)$ converges, and the result follows.