CHAPTER 2

Axioms of Probability

2.1 INTRODUCTION

In this chapter we introduce the concept of the probability of an event and then show how these probabilities can be computed in certain situations. As a preliminary, however, we need the concept of the sample space and the events of an experiment.

2.2 SAMPLE SPACE AND EVENTS

Consider an experiment whose outcome is not predictable with certainty in advance. However, although the outcome of the experiment will not be known in advance, let us suppose that the set of all possible outcomes is known. This set of all possible outcomes of an experiment is known as the *sample space* of the experiment and is denoted by S. Some examples follow.

1. If the outcome of an experiment consists in the determination of the sex of a newborn child, then

$$S = \{g, b\}$$

where the outcome g means that the child is a girl and b that it is a boy.

2. If the outcome of an experiment is the order of finish in a race among the 7 horses having post positions 1, 2, 3, 4, 5, 6, 7, then

$$S = \{\text{all } 7! \text{ permutations of } (1, 2, 3, 4, 5, 6, 7)\}$$

The outcome (2, 3, 1, 6, 5, 4, 7) means, for instance, that the number 2 horse comes in first, then the number 3 horse, then the number 1 horse, and so on.

3. If the experiment consists of flipping two coins, then the sample space consists of the following four points:

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

The outcome will be (H, H) if both coins are heads, (H, T) if the first coin is heads and the second tails, (T, H) if the first is tails and the second heads, and (T, T) if both coins are tails.

4. If the experiment consists of tossing two dice, then the sample space consists of the 36 points

$$S = \{(i, j): i, j = 1, 2, 3, 4, 5, 6\}$$

where the outcome (i, j) is said to occur if i appears on the leftmost die and j on the other die.

5. If the experiment consists of measuring (in hours) the lifetime of a transistor, then the sample space consists of all nonnegative real numbers. That is

$$S = \{x: 0 \le x < \infty\}$$

Any subset E of the sample space is known as an *event*. That is, an event is a set consisting of possible outcomes of the experiment. If the outcome of the experiment is contained in E, then we say that E has occurred. Some examples of events are the following.

In example 1 above, if $E = \{g\}$, then E is the event that the child is a girl. Similarly, if $F = \{b\}$, then F is the event that the child is a boy.

In example 2, if

$$E = \{\text{all outcomes in } S \text{ starting with a } 3\}$$

then E is the event that horse 3 wins the race.

In example 3, if $E = \{(H, H), (H, T)\}$, then E is the event that a head appears on the first coin.

In example 4, if $E = \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}$, then E is the event that the sum of the dice equals 7.

In example 5, if $E = \{x: 0 \le x \le 5\}$, then E is the event that the transistor does not last longer than 5 hours.

For any two events E and F of a sample space S, we define the new event $E \cup F$ to consist of all points that are either in E or in F or in both E and F. That is, the event $E \cup F$ will occur if *either* E or F occurs. For instance, in example 1 if event $E = \{g\}$ and $F = \{b\}$, then

$$E \cup F = \{g, b\}$$

That is, $E \cup F$ would be the whole sample space S. In example 3, if $E = \{(H, H), (H, T)\}$ and $F = \{(T, H)\}$, then

$$E \cup F = \{(H, H), (H, T), (T, H)\}$$

Thus $E \cup F$ would occur if a head appeared on either coin.

The event $E \cup F$ is called the *union* of the event E and the event F.

Similarly, for any two events E and F we may also define the new event EF, called the *intersection* of E and F, to consist of all outcomes that are both in E and in F. That is, the event EF (sometimes written $E \cap F$) will occur only if both E and F occur. For instance, in example 3 if $E = \{(H, H), (H, T),$

$(T, H)\}$ is the event that at least 1 head occurs, and $F = \{(H, T), (T, H), (T, T)\}$ is the event that at least 1 tail occurs, then

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

is the event that exactly 1 head and 1 tail appear. In example 4 if $E = \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}$ is the event that the sum of the dice is 7 and $F = \{(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)\}$ is the event that the sum is 6, then the event EF does not contain any outcomes and hence could not occur. To give such an event a name, we shall refer to it as the null event and denote it by $\varnothing$ (that is, $\varnothing$ refers to the event consisting of no points). If $EF = \varnothing$, then E and F are said to be *mutually exclusive*.

We also define unions and intersections of more than two events in a similar manner. If $E_1, E_2, \ldots$ are events, the union of these events, denoted by $\bigcup_{n=1}^{\infty} E_n$, is defined to be that event which consists of all points that are in E_n for at least one value of $n = 1, 2, \ldots$. Similarly, the intersection of the events E_n, denoted by $\bigcap_{n=1}^{\infty} E_n$, is defined to be the event consisting of those points that are in all of the events E_n, $n = 1, 2, \ldots$.

Finally, for any event E we define the new event E^c, referred to as the complement of E, to consist of all points in the sample space S that are not in E. That is, E^c will occur if and only if E does not occur. In example 4, if event $E = \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}$, then E^c will occur when the sum of the dice does not equal 7. Also note that because the experiment must result in some outcome, it follows that $S^c = \varnothing$.

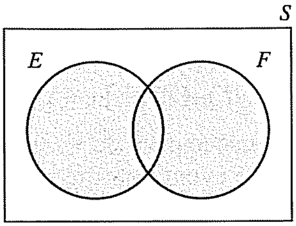
For any two events E and F, if all of the points in E are also in F, then we say that E is contained in F and write $E \subset F$ (or equivalently, $F \supset E$). Thus, if $E \subset F$, the occurrence of E necessarily implies the occurrence of F. If $E \subset F$ and $F \subset E$, we say that E and F are equal and write $E = F$.

A graphical representation that is very useful for illustrating logical relations among events is the Venn diagram. The sample space S is represented as consisting of all the points in a large rectangle, and the events $E, F, G, \ldots$ are represented as consisting of all the points in given circles within the rectangle. Events of interest can then be indicated by shading appropriate regions of the diagram. For instance, in the three Venn diagrams shown in Figure 2.1, the shaded areas represent, respectively, the events $E \cup F$, EF, and E^c. The Venn diagram in Figure 2.2 indicates that $E \subset F$.

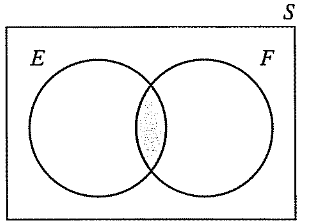
The operations of forming unions, intersections, and complements of events obey certain rules not dissimilar to the rules of algebra. We list a few of these rules.

Commutative laws	$E \cup F = F \cup E$	$EF = FE$
Associative laws	$(E \cup F) \cup G = E \cup (F \cup G)$	$(EF)G = E(FG)$
Distributive laws	$(E \cup F)G = EG \cup FG$	$EF \cup G = (E \cup G)(F \cup G)$

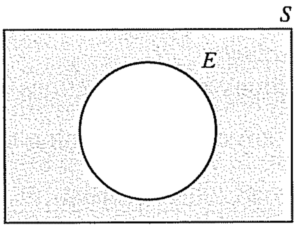
These relations are verified by showing that any outcome that is contained in the event on the left side of the equality sign is also contained in the event on the



(a) Shaded region: $E \cup F$.



(b) Shaded region: EF.



(c) Shaded region: E^c.

Figure 2.1

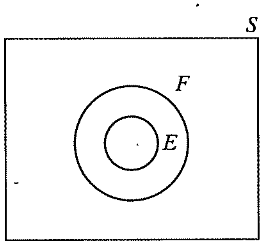
right side, and vice versa. One way of showing this is by means of Venn diagrams. For instance, the distributive law may be verified by the sequence of diagrams in Figure 2.3.

The following useful relationships between the three basic operations of forming unions, intersections, and complements are known as *DeMorgan's laws*:

$$\left(\bigcup_{i=1}^{n} E_i\right)^c = \bigcap_{i=1}^{n} E_i^c$$

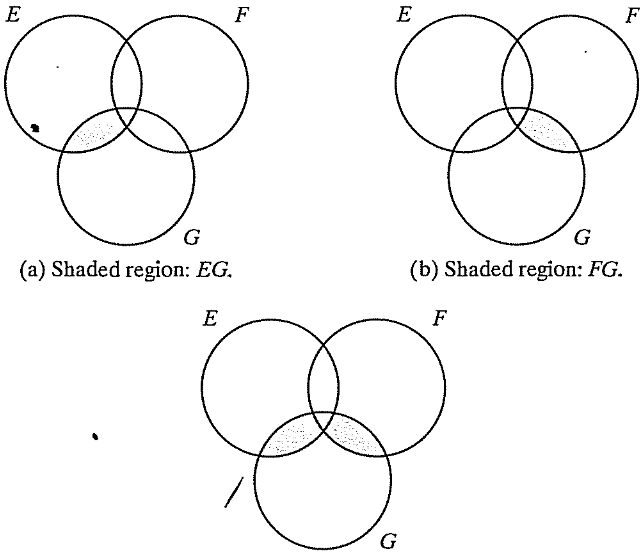
$$\left(\bigcap_{i=1}^{n} E_i\right)^c = \bigcup_{i=1}^{n} E_i^c$$

To prove DeMorgan's laws, suppose first that x is a point of $\left(\bigcup_{i=1}^{n} E_i\right)^c$. Then x is not contained in $\bigcup_{i=1}^{n} E_i$, which means that x is not contained in any of the events E_i, $i = 1, 2, \ldots, n$, implying that x is contained in E_i^c for all $i = 1$,



$E \subset F$

Figure 2.2



(a) Shaded region: EG.

(b) Shaded region: FG.

(c) Shaded region: $(E \cup F)G$.
$(E \cup F)G = EG \cup FG$

Figure 2.3

2, ..., n and thus is contained in $\bigcap_{i=1}^{n} E_i^c$. To go the other way, suppose that x is a point of $\bigcap_{i=1}^{n} E_i^c$. Then x is contained in E_i^c for all $i = 1, 2, \ldots, n$, which means that x is not contained in E_i for any $i = 1, 2, \ldots, n$, implying that x is not contained in $\bigcup_{i}^{n} E_i$, which yields that x is contained in $\left(\bigcup_{1}^{n} E_i\right)^c$. This proves the first of DeMorgan's laws.

To prove the second of DeMorgan's laws, we use the first law to obtain

$$\left(\bigcup_{i=1}^{n} E_i^c\right)^c = \bigcap_{i=1}^{n} (E_i^c)^c$$

which, since $(E^c)^c = E$, is equivalent to

$$\left(\bigcup_{1}^{n} E_i^c\right)^c = \bigcap_{1}^{n} E_i$$

Taking complements of both sides of the above yields the result, namely,

$$\bigcup_{1}^{n} E_i^c = \left(\bigcap_{1}^{n} E_i\right)^c$$

2.3 AXIOMS OF PROBABILITY

One way of defining the probability of an event is in terms of its relative frequency. Such a definition usually goes as follows: We suppose that an experiment, whose sample space is S, is repeatedly performed under exactly the same conditions. For each event E of the sample space S, we define $n(E)$ to be the number of times in the first n repetitions of the experiment that the event E occurs. Then $P(E)$, the probability of the event E, is defined by

$$P(E) = \lim_{n \to \infty} \frac{n(E)}{n}$$

That is, $P(E)$ is defined as the (limiting) proportion of time that E occurs. It is thus the limiting frequency of E.

Although the preceding definition is certainly intuitively pleasing and should always be kept in mind by the reader, it possesses a serious drawback: How do we know that $n(E)/n$ will converge to some constant limiting value that will be the same for each possible sequence of repetitions of the experiment? For example, suppose that the experiment to be repeatedly performed consists of flipping a coin. How do we know that the proportion of heads obtained in the first n flips will converge to some value as n gets large? Also, even if it does converge to some value, how do we know that, if the experiment is repeatedly performed a second time, we shall again obtain the same limiting proportion of heads?

Proponents of the relative frequency definition of probability usually answer this objection by stating that the convergence of $n(E)/n$ to a constant limiting value is an assumption, or an *axiom*, of the system. However, to assume that $n(E)/n$ will necessarily converge to some constant value seems to be a very complex assumption. For, although we might indeed hope that such a constant limiting frequency exists, it does not at all seem to be a priori evident that this need be the case. In fact, would it not be more reasonable to assume a set of simpler and more self-evident axioms about probability and then attempt to prove that such a constant limiting frequency does in some sense exist? This latter approach is the modern axiomatic approach to probability theory that we shall adopt in this text. In particular, we shall assume that for each event E in the sample space S there exists a value $P(E)$, referred to as the probability of E. We shall then assume that the probabilities satisfy a certain set of axioms, which, we hope the reader will agree, is in accordance with our intuitive notion of probability.

Consider an experiment whose sample space is S. For each event E of the sample space S we assume that a number $P(E)$ is defined and satisfies the following three axioms.

Axiom 1

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

Axiom 2

$$P(S) = 1$$

Axiom 3

For any sequence of mutually exclusive events $E_1, E_2, \ldots$ (that is, events for which $E_iE_j = \emptyset$ when $i \neq j$),

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$$

We refer to $P(E)$ as the probability of the event E.

Thus Axiom 1 states that the probability that the outcome of the experiment is a point in E is some number between 0 and 1. Axiom 2 states that, with probability 1, the outcome will be a point in the sample space S. Axiom 3 states that for any sequence of mutually exclusive events the probability of at least one of these events occurring is just the sum of their respective probabilities.

If we consider a sequence of events $E_1, E_2, \ldots$, where $E_1 = S$, $E_i = \emptyset$ for $i > 1$, then, as the events are mutually exclusive and as $S = \bigcup_{i=1}^{\infty} E_i$, we have from Axiom 3 that

$$P(S) = \sum_{i=1}^{\infty} P(E_i) = P(S) + \sum_{i=2}^{\infty} P(\emptyset)$$

implying that

$$P(\emptyset) = 0$$

That is, the null event has probability 0 of occurring.

It should also be noted that it follows that for any finite sequence of mutually exclusive events $E_1, E_2, \ldots, E_n$,

$$P\left(\bigcup_{1}^{n} E_i\right) = \sum_{i=1}^{n} P(E_i) \tag{3.1}$$

This follows from Axiom 3 by defining E_i to be the null event for all values of i greater than n. Axiom 3 is equivalent to Equation (3.1) when the sample space is finite (why?). However, the added generality of Axiom 3 is necessary when the sample space consists of an infinite number of points.

Example 3a. If our experiment consists of tossing a coin and if we assume that a head is as likely to appear as a tail, then we would have

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

On the other hand, if the coin were biased and we felt that a head were twice as likely to appear as a tail, then we would have

$$P(\{H\}) = \tfrac{2}{3} \qquad P(\{T\}) = \tfrac{1}{3}$$

Example 3b. If a die is rolled and we suppose that all six sides are equally likely to appear, then we would have $P(\{1\}) = P(\{2\}) = P(\{3\}) = P(\{4\}) = P(\{5\}) = P(\{6\}) = \frac{1}{6}$. From Axiom 3 it would thus follow that the probability of rolling an even number would equal

$$P(\{2, 4, 6\}) = P(\{2\}) + P(\{4\}) + P(\{6\}) = \tfrac{1}{2}$$

The assumption of the existence of a set function P, defined on the events of a sample space S, and satisfying Axioms 1, 2, and 3, constitutes the modern mathematical approach to probability theory. Hopefully, the reader will agree that the axioms are natural and in accordance with our intuitive concept of probability as related to chance and randomness. Furthermore, using these axioms we shall be able to prove that if an experiment is repeated over and over again then, with probability 1, the proportion of time during which any specific event E occurs will equal $P(E)$. This result, known as the strong law of large numbers, is presented in Chapter 8. In addition, we present another possible interpretation of probability—as being a measure of belief—in Section 2.7.

TECHNICAL REMARK. We have supposed that $P(E)$ is defined for all the events E of the sample space. Actually, when the sample space is an uncountably infinite set $P(E)$ is defined only for a class of events called measurable. However, this restriction need not concern us as all events of any practical interest are measurable.

2.4 SOME SIMPLE PROPOSITIONS

In this section we prove some simple propositions regarding probabilities. We first note that as E and E^c are always mutually exclusive and since $E \cup E^c = S$, we have by Axioms 2 and 3 that

$$1 = P(S) = P(E \cup E^c) = P(E) + P(E^c)$$

Or equivalently, we have the statement given in Proposition 4.1.

Proposition 4.1

$$P(E^c) = 1 - P(E)$$

In words, Proposition 4.1 states that the probability that an event does not occur is 1 minus the probability that it does occur. For instance, if the probability of obtaining a head on the toss of a coin is $\frac{3}{8}$, the probability of obtaining a tail must be $\frac{5}{8}$.

Our second proposition states that if the event E is contained in the event F, then the probability of E is no greater than the probability of F.

Proposition 4.2

If $E \subset F$, then $P(E) \leq P(F)$.

Proof: Since $E \subset F$, it follows that we can express F as

$$F = E \cup E^c F$$

Hence, as E and $E^c F$ are mutually exclusive, we obtain from Axiom 3 that

$$P(F) = P(E) + P(E^c F)$$

which proves the result, since $P(E^c F) \geq 0$.

Proposition 4.2 tells us, for instance, that the probability of rolling a 1 with a die is less than or equal to the probability of rolling an odd value with the die.

The next proposition gives the relationship between the probability of the union of two events in terms of the individual probabilities and the probability of the intersection.

Proposition 4.3

$$P(E \cup F) = P(E) + P(F) - P(EF)$$

Proof: To derive a formula for $P(E \cup F)$, we first note that $E \cup F$ can be written as the union of the two disjoint events E and $E^c F$. Thus from Axiom 3 we obtain that

$$\begin{aligned} P(E \cup F) &= P(E \cup E^c F) \\ &= P(E) + P(E^c F) \end{aligned}$$

Furthermore, since $F = EF \cup E^c F$, we again obtain from Axiom 3 that

$$P(F) = P(EF) + P(E^c F)$$

or, equivalently,

$$P(E^c F) = P(F) - P(EF)$$

thus completing the proof.

Proposition 4.3 could also have been proved by making use of the Venn diagram in Figure 2.4.

Let us divide $E \cup F$ into three mutually exclusive sections, as shown in Figure 2.5. In words, section I represents all the points in E that are not in F (that

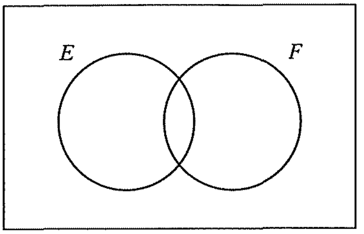


Figure 2.4 Venn diagram.

is, EF^c); section II represents all points both in E and in F (that is, EF); and section III represents all points in F that are not in E (that is, E^cF).

From Figure 2.5 we see that

$$\begin{aligned} E \cup F &= \mathrm{I} \cup \mathrm{II} \cup \mathrm{III} \\ E &= \mathrm{I} \cup \mathrm{II} \\ F &= \mathrm{II} \cup \mathrm{III} \end{aligned}$$

As I, II, and III are mutually exclusive, it follows from Axiom 3 that

$$\begin{aligned} P(E \cup F) &= P(\mathrm{I}) + P(\mathrm{II}) + P(\mathrm{III}) \\ P(E) &= P(\mathrm{I}) + P(\mathrm{II}) \\ P(F) &= P(\mathrm{II}) + P(\mathrm{III}) \end{aligned}$$

which shows that

$$P(E \cup F) = P(E) + P(F) - P(\mathrm{II})$$

and Proposition 4.3 is proved, since $\mathrm{II} = EF$.

Example 4a. Suppose that we toss two coins and suppose that each of the four points in the sample space $S = \{(H, H), (H, T), (T, H), (T, T)\}$ is equally likely and hence has probability $\frac{1}{4}$. Let

$$E = \{(H, H), (H, T)\} \qquad \text{and} \qquad F = \{(H, H), (T, H)\}$$

That is, E is the event that the first coin falls heads, and F is the event that the second coin falls heads.

By Proposition 4.3 we have that $P(E \cup F)$, the probability that either the first or second coin falls heads, is given by

$$\begin{aligned} P(E \cup F) &= P(E) + P(F) - P(EF) \\ &= \tfrac{1}{2} + \tfrac{1}{2} - P(\{H, H)\}) \\ &= 1 - \tfrac{1}{4} \\ &= \tfrac{3}{4} \end{aligned}$$

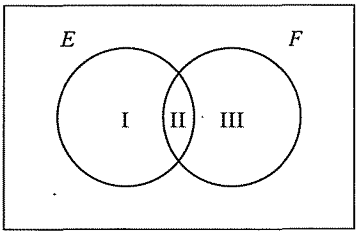


Figure 2.5 Venn diagram in sections.

This probability could, of course, have been computed directly because

$$P(E \cup F) = P(\{(H, H), (H, T), (T, H)\}) = \tfrac{3}{4} \qquad \blacksquare$$

We may also calculate the probability that any one of the three events E or F or G occurs:

$$P(E \cup F \cup G) = P[(E \cup F) \cup G]$$

which by Proposition 4.3 equals

$$P(E \cup F) + P(G) - P[(E \cup F)G]$$

Now, it follows from the distributive law that the events $(E \cup F)G$ and $EG \cup FG$ are equivalent, and hence we obtain from the preceding equations that

$$\begin{aligned} P(E \cup F \cup G) & \\ &= P(E) + P(F) - P(EF) + P(G) - P(EG \cup FG) \\ &= P(E) + P(F) - P(EF) + P(G) - P(EG) - P(FG) + P(EGFG) \\ &= P(E) + P(F) + P(G) - P(EF) - P(EG) - P(FG) + P(EFG) \end{aligned}$$

In fact, the following proposition can be proved by induction.

Proposition 4.4

$$\begin{aligned} P(E_1 \cup E_2 \cup \cdots \cup E_n) = \sum_{i=1}^{n} P(E_i) - \sum_{i_1<i_2} P(E_{i_1}E_{i_2}) + \cdots \\ + (-1)^{r+1} \sum_{i_1<i_2<\cdots<i_r} P(E_{i_1}E_{i_2}\cdots E_{i_r}) \\ + \cdots + (-1)^{n+1} P(E_1E_2\cdots E_n) \end{aligned}$$

The summation $\sum_{i_1<i_2<\cdots<i_r} P(E_{i_1}E_{i_2}\cdots E_{i_r})$ is taken over all of the $\binom{n}{r}$ possible subsets of size r of the set $\{1, 2, \ldots, n\}$.

In words, Proposition 4.4 states that the probability of the union of n events equals the sum of the probabilities of these events taken one at a time, minus the sum of the probabilities of these events taken two at a time, plus the sum of the probabilities of these events taken three at a time, and so on.

REMARK. For a noninductive argument for Proposition 4.4, note first that if a point of the sample space is not a member of any of the sets E_i then its probability does not contribute anything to either side of the equality. On the other hand, suppose that a point is in exactly m of the events E_i, where $m > 0$. Then since it is in $\bigcup_i E_i$ its probability is counted once in $P\left(\bigcup_i E_i\right)$; also as this point is contained in $\binom{m}{k}$ subsets of the type $E_{i_1}E_{i_2}\cdots E_{i_k}$, its probability is counted

$$\binom{m}{1}-\binom{m}{2}+\binom{m}{3}-\cdots\pm\binom{m}{m}$$

times on the right of the equality sign in Proposition 4.4. Thus, for $m > 0$, we must show that

$$1=\binom{m}{1}-\binom{m}{2}+\binom{m}{3}-\cdots\pm\binom{m}{m}$$

However, since $1=\binom{m}{0}$, the preceding is equivalent to

$$\sum_{i=0}^{m}\binom{m}{i}(-1)^i=0$$

and the latter equation follows from the binomial theorem since

$$0=(-1+1)^m=\sum_{i=0}^{m}\binom{m}{i}(-1)^i(1)^{m-i}$$

2.5 SAMPLE SPACES HAVING EQUALLY LIKELY OUTCOMES

For many experiments it is natural to assume that all outcomes in the sample space are equally likely to occur. That is, consider an experiment whose sample space S is a finite set, say $S=\{1,2,\ldots,N\}$. Then it is often natural to assume that

$$P(\{1\})=P(\{2\})=\cdots=P(\{N\})$$

which implies from Axioms 2 and 3 (why?) that

$$P(\{i\})=\frac{1}{N}\qquad i=1,2,\ldots,N$$

From this it follows from Axiom 3 that for any event E

$$P(E)=\frac{\text{number of points in } E}{\text{number of points in } S}$$

In words, if we assume that all outcomes of an experiment are equally likely to occur, then the probability of any event E equals the proportion of points in the sample space that are contained in E.

Example 5a. If two dice are rolled, what is the probability that the sum of the upturned faces will equal 7?

Solution We shall solve this problem under the assumption that all of the 36 possible outcomes are equally likely. Since there are 6 possible outcomes, namely (1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1), that result in the sum of the dice being equal to 7, the desired probability is $\frac{6}{36}=\frac{1}{6}$. ∎

Example 5b. If 3 balls are "randomly drawn" from a bowl containing 6 white and 5 black balls, what is the probability that one of the drawn balls is white and the other two black?

Solution If we regard the order in which the balls are selected as being relevant, then the sample space consists of $11 \cdot 10 \cdot 9 = 990$ outcomes. Furthermore, there are $6 \cdot 5 \cdot 4 = 120$ outcomes in which the first ball selected is white and the other two black, $5 \cdot 6 \cdot 4 = 120$ outcomes in which the first is black, the second white, and the third black; and $5 \cdot 4 \cdot 6 = 120$ in which the first two are black and the third white. Hence, assuming that "randomly drawn" means that each outcome in the sample space is equally likely to occur, we see that the desired probability is

$$\frac{120 + 120 + 120}{990} = \frac{4}{11}$$

This problem could also have been solved by regarding the outcome of the experiment as the unordered set of drawn balls. From this point of view, there are $\binom{11}{3} = 165$ outcomes in the sample space. Now, each set of 3 balls corresponds to $3!$ outcomes when the order of selection is noted. As a result, if all outcomes are assumed equally likely when the order of selection is noted, then it follows that they remain equally likely when the outcome is taken to be the unordered set of selected balls. Hence, using the latter representation of the experiment, we see that the desired probability is

$$\frac{\binom{6}{1}\binom{5}{2}}{\binom{11}{3}} = \frac{4}{11}$$

which, of course, agrees with the answer obtained previously. ▮

Example 5c. A committee of 5 is to be selected from a group of 6 men and 9 women. If the selection is made randomly, what is the probability that the committee consists of 3 men and 2 women?

Solution Let us assume that *randomly selected* means that each of the $\binom{15}{5}$ possible combinations is equally likely to be selected. Hence the desired probability equals

$$\frac{\binom{6}{3}\binom{9}{2}}{\binom{15}{5}} = \frac{240}{1001}$$ ▮

Example 5d. An urn contains n balls, of which one is special. If k of these balls are withdrawn one at a time, with each selection being equally likely to be any of the balls that remain at the time, what is the probability that the special ball is chosen?

Solution Since all of the balls are treated in an identical manner, it follows that the set of k balls selected is equally likely to be any of the $\binom{n}{k}$ sets of k balls. Therefore,

$$P\{\text{special ball is selected}\} = \frac{\binom{1}{1}\binom{n-1}{k-1}}{\binom{n}{k}} = \frac{k}{n}$$

We could also have obtained the preceding result by letting A_i denote the event that the special ball is the ith ball to be chosen, $i = 1, \ldots, k$. Then, since each one of the n balls is equally likely to be the ith ball chosen, it follows that $P(A_i) = 1/n$. Hence, since these events are obviously mutually exclusive, we have that

$$P\{\text{special ball is selected}\} = P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) = \frac{k}{n}$$

We could have argued that $P(A_i) = 1/n$, by noting that there are $n(n-1)\cdots(n-k+1) = n!/(n-k)!$ equally likely outcomes of the experiment, of which $(n-1)(n-2)\cdots(n-i+1)(1)(n-i)\cdots(n-k+1) = (n-1)!/(n-k)!$ result in the special ball being the ith one chosen. From this it follows that

$$P(A_i) = \frac{(n-1)!}{n!} = \frac{1}{n}$$ ∎

Example 5e. Suppose that $n + m$ balls, of which n are red and m are blue, are arranged in a linear order in such a way that all $(n + m)!$ possible orderings are equally likely. If we record the result of this experiment by only listing the colors of the successive balls, show that all the possible results remain equally likely.

Solution Consider any one of the $(n + m)!$ possible orderings and note that any permutation of the red balls among themselves and of the blue balls among themselves does not change the sequence of colors. As a result, every ordering of colorings corresponds to $n!\,m!$ different orderings of the $n + m$ balls, so every ordering of the colors has probability $\dfrac{n!\,m!}{(n+m)!}$ of occurring.

For example, suppose that there are 2 red balls, numbered r_1, r_2 and 2 blue balls, numbered b_1, b_2. Then, of the 4! possible orderings, there will be 2! 2! orderings that result in any specified color combination. For instance, the following orderings result in the successive balls alternating in color with a red ball first:

$$r_1, b_1, r_2, b_2 \qquad r_1, b_2, r_2, b_1 \qquad r_2, b_1, r_1, b_2 \qquad r_2, b_2, r_1, b_1$$

Hence each of the possible orderings of the colors has probability $\frac{4}{24} = \frac{1}{6}$ of occurring. ∎

Example 5f. A poker hand consists of 5 cards. If the cards have distinct consecutive values and are not all of the same suit, we say that the hand is a straight. For instance, a hand consisting of the five of spades, six of spades, seven of spades, eight of spades, and nine of hearts is a straight. What is the probability that one is dealt a straight?

Solution We start by assuming that all $\binom{52}{5}$ possible poker hands are equally likely. To determine the number of outcomes that are straights, let us first determine the number of possible outcomes for which the poker hand consists of an ace, two, three, four, and five (the suits being irrelevant). Since the ace can be any 1 of the 4 possible aces, and similarly for the two, three, four, and five, it follows that there are 4^5 outcomes leading to exactly one ace, two, three, four, and five. Hence, since in 4 of these outcomes all the cards will be of the same suit (such a hand is called a straight flush), it follows that there are $4^5 - 4$ hands that make up a straight of the form ace, two, three, four, and five. Similarly, there are $4^5 - 4$ hands that make up a straight of the form ten, jack, queen, king, and ace. Hence there are $10(4^5 - 4)$ hands that are straights. Thus the desired probability is

$$\frac{10(4^5 - 4)}{\binom{52}{5}} \approx .0039$$ ∎

Example 5g. A 5-card poker hand is said to be a full house if it consists of 3 cards of the same denomination and 2 cards of the same denomination. (That is, a full house is three of a kind plus a pair.) What is the probability that one is dealt a full house?

Solution Again we assume that all $\binom{52}{5}$ possible hands are equally likely. To determine the number of possible full houses, we first note that there are $\binom{4}{2}\binom{4}{3}$ different combinations of, say, 2 tens and 3 jacks. Because there are 13 different choices for the kind of pair and, after a pair has been chosen, there are 12 other choices for the denomination of the remaining 3 cards, it follows that the probability of a full house is

$$\frac{13 \cdot 12 \cdot \binom{4}{2}\binom{4}{3}}{\binom{52}{5}} \approx .0014$$ ∎

CHAPTER 3

Conditional Probability and Independence

3.1 INTRODUCTION

In this chapter we introduce one of the most important concepts in probability theory, that of conditional probability. The importance of this concept is twofold. In the first place, we are often interested in calculating probabilities when some partial information concerning the result of the experiment is available; in such a situation the desired probabilities are conditional. Second, even when no partial information is available, conditional probabilities can often be used to compute the desired probabilities more easily.

3.2 CONDITIONAL PROBABILITIES

Suppose that we toss 2 dice and suppose that each of the 36 possible outcomes is equally likely to occur and hence has probability $\frac{1}{36}$. Suppose further that we observe that the first die is a 3. Then, given this information, what is the probability that the sum of the 2 dice equals 8? To calculate this probability, we reason as follows: Given that the initial die is a 3, it follows that there can be at most 6 possible outcomes of our experiment, namely, (3, 1), (3, 2), (3, 3), (3, 4), (3, 5), and (3, 6). Since each of these outcomes originally had the same probability of occurring, the outcomes should still have equal probabilities. That is, given that the first die is a 3, the (conditional) probability of each of the outcomes (3, 1), (3, 2), (3, 3), (3, 4), (3, 5), and (3, 6) is $\frac{1}{6}$, whereas the (conditional) probability of the other 30 points in the sample space is 0. Hence the desired probability will be $\frac{1}{6}$.

If we let E and F denote, respectively, the event that the sum of the dice is 8 and the event that the first die is a 3, then the probability just obtained is called the conditional probability that E occurs given that F has occurred and is denoted by

$$P(E|F)$$

A general formula for $P(E|F)$ that is valid for all events E and F is derived in the same manner: If the event F occurs, then in order for E to occur it is necessary that the actual occurrence be a point in both E and in F; that is, it must be in EF. Now, as we know that F has occurred, it follows that F becomes our new or reduced sample space; hence the probability that the event EF occurs will equal the probability of EF relative to the probability of F. That is, we have the following definition.

Definition

If $P(F) > 0$, then

$$P(E|F) = \frac{P(EF)}{P(F)} \tag{2.1}$$

Example 2a. A coin is flipped twice. If we assume that all four points in the sample space $S = \{(H, H), (H, T), (T, H), (T, T)\}$ are equally likely, what is the conditional probability that both flips result in heads, given that the first flip does?

Solution If $E = \{(H, H)\}$ denotes the event that both flips land heads, and $F = \{(H, H), (H, T)\}$ the event that the first flip lands heads, then the desired probability is given by

$$\begin{aligned} P(E|F) &= \frac{P(EF)}{P(F)} \\ &= \frac{P(\{(H, H)\})}{P(\{(H, H), (H, T)\})} \\ &= \frac{\frac{1}{4}}{\frac{2}{4}} = \frac{1}{2} \end{aligned}$$

∎

Example 2b. An urn contains 10 white, 5 yellow, and 10 black marbles. A marble is chosen at random from the urn, and it is noted that it is not one of the black marbles. What is the probability that it is yellow?

Solution Let Y denote the event that the marble selected is yellow, and let B^c denote the event that it is not black. Now, from Equation (2.1),

$$P(Y|B^c) = \frac{P(YB^c)}{P(B^c)}$$

However, $YB^c = Y$, since the marble will be both yellow and not black if and only if it is yellow. Hence, assuming that each of the 25 marbles is equally likely to be chosen, we obtain that

$$P(Y|B^c) = \frac{\frac{5}{25}}{\frac{15}{25}} = \frac{1}{3}$$

It should be noted that we also could have derived this probability by working directly with the reduced sample space. That is, as we know that the chosen marble is not black, the problem reduces to computing the probability that a marble, chosen at random from an urn containing 10 white and 5 yellow marbles, is yellow. This is clearly equal to $\frac{5}{15} = \frac{1}{3}$. ∎

When all outcomes are assumed to be equally likely, it is often easier to compute a conditional probability by a consideration of the reduced sample space, as opposed to a direct application of (2.1).

Example 2c. In the card game bridge the 52 cards are dealt out equally to 4 players—called East, West, North, and South. If North and South have a total of 8 spades among them, what is the probability that East has 3 of the remaining 5 spades?

Solution Probably the easiest way to compute this is to work with the reduced sample space. That is, given that North–South have a total of 8 spades among their 26 cards, there remains a total of 26 cards, exactly 5 of them being spades, to be distributed among the East–West hands. As each distribution is equally likely, it follows that the conditional probability that East will have exactly 3 spades among his or her 13 cards is

$$\frac{\binom{5}{3}\binom{21}{10}}{\binom{26}{13}} \approx .339$$

∎

Example 2d. The organization for which Ms. Jones works is running a dinner for those employees having at least one son. If Jones is known to have two children, what is the conditional probability that they are both boys, given that she is invited to the dinner? Assume that the sample space S is given by $S = \{(b, b), (b, g), (g, b), (g, g)\}$ and all outcomes are equally likely [(b, g) means, for instance, that the older child is a boy and the younger child is a girl].

Solution The knowledge that Jones has been invited to the dinner is equivalent to knowing that she has at least one son. Hence, letting E denote the event that both children are boys and F the event that at least one of them is a boy, we have that the desired probability $P(E|F)$ is given by

$$\begin{aligned} P(E|F) &= \frac{P(EF)}{P(F)} \\ &= \frac{P(\{(b, b)\})}{P(\{(b, b), (b, g), (g, b)\})} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3} \end{aligned}$$

Many readers incorrectly reason that the conditional probability of two boys given at least one is $\frac{1}{2}$, as opposed to the correct $\frac{1}{3}$, since they reason that the Jones child not attending the dinner is equally likely to be a boy or a girl. Their mistake, however, is in assuming that these two possibilities are equally likely. For, initially, there were 4 equally likely outcomes. Now the information that at least one child is a boy is equivalent to knowing that the outcome is not (g, g). Hence we are left with the 3 equally likely outcomes (b, b), (b, g), (g, b) thus showing that the Jones child not attending the dinner is twice as likely to be a girl as to be a boy. ∎

By multiplying both sides of Equation (2.1) by $P(F)$, we obtain

$$P(EF) = P(F)P(E|F) \tag{2.2}$$

In words, Equation (2.2) states that the probability that both E and F occur is equal to the probability that F occurs multiplied by the conditional probability of E given that F occurred. Equation (2.2) is often quite useful in computing the probability of the intersection of events.

Example 2e. Celine is undecided as to whether to take a French course or a chemistry course. She estimates that her probability of receiving an A grade would be $\frac{1}{2}$ in a French course, and $\frac{2}{3}$ in a chemistry course. If Celine decides to base her decision on the flip of a fair coin, what is the probability that she gets an A in chemistry?

Solution If we let C be the event that Celine takes chemistry and A denote the event that she receives an A in whatever course she takes, then the desired probability is $P(CA)$. This is calculated by using Equation (2.2) as follows:

$$\begin{aligned} P(CA) &= P(C)P(A|C) \\ &= \left(\tfrac{1}{2}\right)\left(\tfrac{2}{3}\right) = \tfrac{1}{3} \end{aligned}$$

∎

Example 2f. Suppose that an urn contains 8 red balls and 4 white balls. We draw 2 balls from the urn without replacement. If we assume that at each draw each ball in the urn is equally likely to be chosen, what is the probability that both balls drawn are red?

Solution Let R_1 and R_2 denote, respectively, the events that the first and second balls drawn are red. Now, given that the first ball selected is red, there are 7 remaining red balls and 4 white balls, and so $P(R_2|R_1) = \frac{7}{11}$. As $P(R_1)$ is clearly $\frac{8}{12}$, the desired probability is

$$\begin{aligned} P(R_1R_2) &= P(R_1)P(R_2|R_1) \\ &= \left(\tfrac{2}{3}\right)\left(\tfrac{7}{11}\right) = \tfrac{14}{33} \end{aligned}$$

Of course, this probability could also have been computed by

$$P(R_1R_2) = \frac{\binom{8}{2}}{\binom{12}{2}}$$

∎

A generalization of Equation (2.2), which provides an expression for the probability of the intersection of an arbitrary number of events, is sometimes referred to as the *multiplication rule.*

The multiplication rule

$$P(E_1E_2E_3\cdots E_n) = P(E_1)P(E_2|E_1)P(E_3|E_1E_2)\cdots P(E_n|E_1\cdots E_{n-1})$$

To prove the multiplication rule, just apply the definition of conditional probability to its right-hand side. This gives

$$P(E_1)\frac{P(E_1E_2)}{P(E_1)}\frac{P(E_1E_2E_3)}{P(E_1E_2)}\cdots\frac{P(E_1E_2\cdots E_n)}{P(E_1E_2\cdots E_{n-1})} = P(E_1E_2\cdots E_n)$$

We will now employ the multiplication rule to obtain a second approach to solving Example 5h(b) of Chapter 2.

Example 2g. An ordinary deck of 52 playing cards is randomly divided into 4 piles of 13 cards each. Compute the probability that each pile has exactly 1 ace.

Solution Define events E_i, $i = 1, 2, 3, 4$ as follows:

$$\begin{aligned}
E_1 &= \{\text{the ace of spades is in any one of the piles}\}\\
E_2 &= \{\text{the ace of spades and the ace of hearts}\\
&\qquad \text{are in different piles}\}\\
E_3 &= \{\text{the aces of spades, hearts, and diamonds}\\
&\qquad \text{are all in different piles}\}\\
E_4 &= \{\text{all 4 aces are in different piles}\}
\end{aligned}$$

The probability desired is $P(E_1E_2E_3E_4)$ and by the multiplication rule

$$P(E_1E_2E_3E_4) = P(E_1)P(E_2|E_1)P(E_3|E_1E_2)P(E_4|E_1E_2E_3)$$

Now,

$$P(E_1) = 1$$

since E_1 is the sample space S.

$$P(E_2|E_1) = \tfrac{39}{51}$$

since the pile containing the ace of spades will receive 12 of the remaining 51 cards.

$$P(E_3|E_1E_2) = \tfrac{26}{50}$$

since the piles containing the aces of spades and hearts will receive 24 of the remaining 50 cards; and finally,

$$P(E_4|E_1E_2E_3) = \tfrac{13}{49}$$

Therefore, we obtain that the probability that each pile has exactly 1 ace is

$$P(E_1E_2E_3E_4) = \frac{39 \cdot 26 \cdot 13}{51 \cdot 50 \cdot 49} \approx .105$$

That is, there is approximately a 10.5 percent chance that each pile will contain an ace. (Problem 20 gives another way of using the multiplication rule to solve this problem.) ∎

REMARK. Our definition of $P(E|F)$ is consistent with the interpretation of probability as being a long-run relative frequency. To see this, suppose that n repetitions of the experiment are to be performed, where n is large. We claim that if we consider only those experiments in which F occurs, then $P(E|F)$ will equal the long-run proportion of them in which E also occurs. To verify this, note that since $P(F)$ is the long-run proportion of experiments in which F occurs, it follows that in the n repetitions of the experiment F will occur approximately $nP(F)$ times. Similarly, in approximately $nP(EF)$ of these experiments both E and F will occur. Hence, out of the approximately $nP(F)$ experiments in which F occurs, the proportion of them in which E also occurs is approximately equal to

$$\frac{nP(EF)}{nP(F)} = \frac{P(EF)}{P(F)}$$

As this approximation becomes exact as n becomes larger and larger, we see that we have the appropriate definition of $P(E|F)$.

3.3 BAYES' FORMULA

Let E and F be events. We may express E as

$$E = EF \cup EF^c$$

for in order for a point to be in E, it must either be in both E and F or be in E but not in F (see Figure 3.1). As EF and EF^c are clearly mutually exclusive, we have by Axiom 3 that

$$\boxed{\begin{aligned} P(E) &= P(EF) + P(EF^c) \\ &= P(E|F)P(F) + P(E|F^c)P(F^c) \\ &= P(E|F)P(F) + P(E|F^c)[1 - P(F)] \end{aligned}} \tag{3.1}$$

Equation (3.1) states that the probability of the event E is a weighted average of the conditional probability of E given that F has occurred and the conditional probability of E given that F has not occurred—each conditional probability being given as much weight as the event on which it is conditioned has of occurring. This is an extremely useful formula because its use often enables us to determine the probability of an event by first "conditioning" upon whether or not some

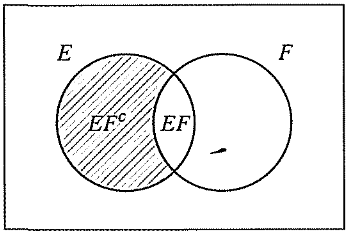


Figure 3.1 $E = EF \cup EF^c$. EF = Shaded Area; EF^c = Striped Area.

second event has occurred. That is, there are many instances where it is difficult to compute the probability of an event directly, but it is straightforward to compute it once we know whether or not some second event has occurred. We illustrate this with some examples.

Example 3a (Part 1). An insurance company believes that people can be divided into two classes: those who are accident prone and those who are not. Their statistics show that an accident-prone person will have an accident at some time within a fixed 1-year period with probability .4, whereas this probability decreases to .2 for a non-accident-prone person. If we assume that 30 percent of the population is accident prone, what is the probability that a new policyholder will have an accident within a year of purchasing a policy?

Solution We shall obtain the desired probability by first conditioning upon whether or not the policyholder is accident prone. Let A_1 denote the event that the policyholder will have an accident within a year of purchase; and let A denote the event that the policyholder is accident prone. Hence the desired probability, $P(A_1)$, is given by

$$\begin{aligned} P(A_1) &= P(A_1|A)P(A) + P(A_1|A^c)P(A^c) \\ &= (.4)(.3) + (.2)(.7) = .26 \end{aligned}$$ ∎

Example 3a (Part 2). Suppose that a new policyholder has an accident within a year of purchasing a policy. What is the probability that he or she is accident prone?

Solution The desired probability is $P(A|A_1)$, which is given by

$$\begin{aligned} P(A|A_1) &= \frac{P(AA_1)}{P(A_1)} \\ &= \frac{P(A)P(A_1|A)}{P(A_1)} \\ &= \frac{(.3)(.4)}{.26} = \frac{6}{13} \end{aligned}$$ ∎

Example 3b. In answering a question on a multiple-choice test, a student either knows the answer or guesses. Let p be the probability that the student knows the answer and $1 - p$ the probability that the student guesses. Assume that a student who guesses at the answer will be correct with probability $1/m$, where m is the number of multiple-choice alternatives. What is the conditional probability that a student knew the answer to a question, given that he or she answered it correctly?

Solution Let C and K denote, respectively, the events that the student answers the question correctly and the event that he or she actually knows the answer. Now

$$\begin{aligned}
P(K|C) &= \frac{P(KC)}{P(C)} \\
&= \frac{P(C|K)P(K)}{P(C|K)P(K) + P(C|K^c)P(K^c)} \\
&= \frac{p}{p + (1/m)(1-p)} \\
&= \frac{mp}{1 + (m-1)p}
\end{aligned}$$

For example, if $m = 5$, $p = \frac{1}{2}$, then the probability that a student knew the answer to a question he or she correctly answered is $\frac{5}{6}$. ∎

Example 3c. A laboratory blood test is 95 percent effective in detecting a certain disease when it is, in fact, present. However, the test also yields a "false positive" result for 1 percent of the healthy persons tested. (That is, if a healthy person is tested, then, with probability .01, the test result will imply he or she has the disease.) If .5 percent of the population actually has the disease, what is the probability a person has the disease given that the test result is positive?

Solution Let D be the event that the tested person has the disease and E the event that the test result is positive. The desired probability $P(D|E)$ is obtained by

$$\begin{aligned}
P(D|E) &= \frac{P(DE)}{P(E)} \\
&= \frac{P(E|D)P(D)}{P(E|D)P(D) + P(E|D^c)P(D^c)} \\
&= \frac{(.95)(.005)}{(.95)(.005) + (.01)(.995)} \\
&= \frac{95}{294} \approx .323
\end{aligned}$$

Thus only 32 percent of those persons whose test results are positive actually have the disease. As many students are often surprised at this result (as they expected this figure to be much higher, since the blood test seems to be a good one), it is probably worthwhile to present a second argument that, although less rigorous than the preceding one, is probably more revealing. We now do so.

Since .5 percent of the population actually has the disease, it follows that, on the average, 1 person out of every 200 tested will have it. The test will correctly confirm that this person has the disease with probability .95. Thus, on the average out of every 200 persons tested, the test will correctly confirm that .95 persons have the disease. On the other hand, however, out of the (on the average) 199 healthy people, the test will incorrectly state that (199)(.01) of these people have the disease. Hence, for every .95 diseased person that the test correctly states is ill, there are (on the average) (199)(.01) healthy persons that the test incorrectly states are ill. Hence the proportion of time that the test result is correct when it states that a person is ill is

$$\frac{.95}{.95 + (199)(.01)} = \frac{95}{294} \approx .323 \qquad \blacksquare$$

Equation (3.1) is also useful when one has to reassess one's personal probabilities in the light of additional information. For instance, consider the following examples.

Example 3d. Consider a medical practitioner pondering the following dilemma: "If I'm at least 80 percent certain that my patient has this disease, then I always recommend surgery, whereas if I'm not quite as certain, then I recommend additional tests that are expensive and sometimes painful. Now, initially I was only 60 percent certain that Jones had the disease, so I ordered the series A test, which always gives a positive result when the patient has the disease and almost never does when he is healthy. The test result was positive, and I was all set to recommend surgery when Jones informed me, for the first time, that he is a diabetic. This information complicates matters because, although it doesn't change my original 60 percent estimate of his chances of having the disease, it does affect the interpretation of the results of the A test. This is so because the A test, while never yielding a positive result when the patient is healthy, does unfortunately yield a positive result 30 percent of the time in the case of *diabetic* patients not suffering from the disease. Now what do I do? More tests or immediate surgery?"

Solution In order to decide whether or not to recommend surgery, the doctor should first compute his updated probability that Jones has the disease given that the A test result was positive. Let D denote the event that Jones has the disease, and E the event of a positive A test result. The desired conditional probability $P(D|E)$ is obtained by

$$\begin{aligned} P(D|E) &= \frac{P(DE)}{P(E)} \\ &= \frac{P(D)P(E|D)}{P(E|D)P(D) + P(E|D^c)P(D^c)} \\ &= \frac{(.6)1}{1(.6) + (.3)(.4)} \\ &= .833 \end{aligned}$$

Note that we have computed the probability of a positive test result by conditioning on whether or not Jones has the disease and then using the fact that because Jones is a diabetic his conditional probability of a positive result given he does not have the disease, $P(E|D^c)$, equals .3. Hence, as the doctor should now be over 80 percent certain that Jones has the disease, he should recommend surgery. ▮

Example 3e. At a certain stage of a criminal investigation the inspector in charge is 60 percent convinced of the guilt of a certain suspect. Suppose now that a *new* piece of evidence that shows that the criminal has a certain characteristic (such as left-handedness, baldness, or brown hair) is uncovered. If 20 percent of the population possesses this characteristic, how certain of the guilt of the suspect should the inspector now be if it turns out that the suspect has this characteristic?

Solution Letting G denote the event that the suspect is guilty and C the event that he possesses the characteristic of the criminal, we have

$$\begin{aligned}
P(G|C) &= \frac{P(GC)}{P(C)} \\
&= \frac{P(C|G)P(G)}{P(C|G)P(G) + P(C|G^c)P(G^c)} \\
&= \frac{1(.6)}{1(.6) + (.2)(.4)} \\
&\approx .882
\end{aligned}$$

where we have supposed that the probability of the suspect having the characteristic if he is, in fact, innocent is equal to .2, the proportion of the population possessing the characteristic. ▮

Example 3f. In the world bridge championships held in Buenos Aires in May 1965 the famous British bridge partnership of Terrence Reese and Boris Schapiro was accused of cheating by using a system of finger signals that could indicate the number of hearts held by the players. Reese and Schapiro denied the accusation, and eventually a hearing was held by the British bridge league. The hearing was in the form of a legal proceeding with a prosecuting and defense team, both having the power to call and cross-examine witnesses. During the course of these proceedings the prosecutor examined specific hands played by Reese and Schapiro and claimed that their playing in these hands was consistent with the hypothesis that they were guilty of having illicit knowledge of the heart suit. At this point, the defense attorney pointed out that their play of these hands was also perfectly consistent with their standard line of play. However, the prosecution then argued that as long as their play was consistent with the hypothesis of guilt, then it must be counted as evidence toward this hypothesis. What do you think of the reasoning of the prosecution?

Solution The problem is basically one of determining how the introduction of new evidence (in the above example, the playing of the hands) affects the probability of a particular hypothesis. Now, if we let H denote a particular hypothesis (such as the guilt of Reese and Schapiro), and E the new evidence, then

$$\begin{aligned} P(H|E) &= \frac{P(HE)}{P(E)} \\ &= \frac{P(E|H)P(H)}{P(E|H)P(H) + P(E|H^c)[1 - P(H)]} \end{aligned} \tag{3.2}$$

where $P(H)$ is our evaluation of the likelihood of the hypothesis before the introduction of the new evidence. The new evidence will be in support of the hypothesis whenever it makes the hypothesis more likely, that is, whenever $P(H|E) \geq P(H)$. From Equation (3.2), this will be the case whenever

$$P(E|H) \geq P(E|H)P(H) + P(E|H^c)[1 - P(H)]$$

or, equivalently, whenever

$$P(E|H) \geq P(E|H^c)$$

In other words, any new evidence can be considered to be in support of a particular hypothesis only if its occurrence is more likely when the hypothesis is true than when it is false. In fact, the new probability of the hypothesis depends on its initial probability and the ratio of these conditional probabilities, since from Equation (3.2),

$$P(H|E) = \frac{P(H)}{P(H) + [1 - P(H)]\dfrac{P(E|H^c)}{P(E|H)}}$$

Hence, in the problem under consideration, the play of the cards can be considered to support the hypothesis of guilt only if such playing would have been more likely if the partnership were cheating than if they were not. As the prosecutor never made this claim, his assertion that the evidence is in support of the guilt hypothesis is invalid. ∎

The change in the probability of a hypothesis when new evidence is introduced can be expressed compactly in terms of the change in the *odds ratio* of this hypothesis, where the concept of odds ratio is defined as follows.

Definition

The odds ratio of an event A is defined by

$$\frac{P(A)}{P(A^c)} = \frac{P(A)}{1 - P(A)}$$

That is, the odds ratio of an event A tells how much more likely it is that the event A occurs than it is that it does not occur. For instance, if $P(A) = \frac{2}{3}$, then $P(A) = 2P(A^c)$, so the odds ratio is 2. If the odds ratio is equal to α, then it is common to say that the odds are "α to 1" in favor of the hypothesis.

Consider now a hypothesis H that is true with probability $P(H)$ and suppose that new evidence E is introduced. Then the conditional probabilities, given the evidence E, that H is true and that H is not true are given by

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)} \qquad P(H^c|E) = \frac{P(E|H^c)P(H^c)}{P(E)}$$

Therefore, the new odds ratio after the evidence E has been introduced is

$$\boxed{\frac{P(H|E)}{P(H^c|E)} = \frac{P(H)}{P(H^c)}\frac{P(E|H)}{P(E|H^c)}} \tag{3.3}$$

That is, the new value of the odds ratio of H is its old value multiplied by the ratio of the conditional probability of the new evidence given that H is true to the conditional probability given that H is not true. This verifies the result of Example 3f, since the odds ratio, and thus the probability of H, increases whenever the new evidence is more likely when H is true than when it is false. Similarly, the odds ratio decreases whenever the new evidence is more likely when H is false than when it is true.

Example 3g. When coin A is flipped it comes up heads with probability $\frac{1}{4}$, whereas when coin B is flipped it comes up heads with probability $\frac{3}{4}$. Suppose that one of these coins is randomly chosen and is flipped twice. If both flips land heads, what is the probability that coin B was the one flipped?

Solution Let B be the event that coin B was the one flipped. Since $P(B) = P(B^c)$, we obtain from Equation (3.3) that

$$\frac{P(B|\text{two heads})}{P(B^c|\text{two heads})} = \frac{\frac{9}{16}}{\frac{1}{16}} = 9$$

Hence the odds are 9:1, or equivalently the probability is $\frac{9}{10}$, that coin B was the one flipped. ∎

Equation (3.1) may be generalized in the following manner: Suppose that $F_1, F_2, \ldots, F_n$ are mutually exclusive events such that

$$\bigcup_{i=1}^{n} F_i = S$$

In other words, exactly one of the events $F_1, F_2, \ldots, F_n$ must occur. By writing

$$E = \bigcup_{i=1}^{n} EF_i$$

and using the fact that the events EF_i, $i = 1, \ldots, n$ are mutually exclusive, we obtain that

$$\begin{aligned} P(E) &= \sum_{i=1}^{n} P(EF_i) \\ &= \sum_{i=1}^{n} P(E|F_i)P(F_i) \end{aligned} \tag{3.4}$$

Thus Equation (3.4) shows how, for given events $F_1, F_2, \ldots, F_n$ of which one and only one must occur, we can compute $P(E)$ by first conditioning on which one of the F_i occurs. That is, Equation (3.4) states that $P(E)$ is equal to a weighted average of $P(E|F_i)$, each term being weighted by the probability of the event on which it is conditioned.

Suppose now that E has occurred and we are interested in determining which one of the F_j also occurred. By Equation (3.4), we have the following proposition.

Proposition 3.1

$$\begin{aligned} P(F_j|E) &= \frac{P(EF_j)}{P(E)} \\ &= \frac{P(E|F_j)P(F_j)}{\sum_{i=1}^{n} P(E|F_i)P(F_i)} \end{aligned} \tag{3.5}$$

Equation (3.5) is known as Bayes' formula, after the English philosopher Thomas Bayes. If we think of the events F_j as being possible "hypotheses" about some subject matter, then Bayes' formula may be interpreted as showing us how opinions about these hypotheses held before the experiment [that is, the $P(F_j)$] should be modified by the evidence of the experiment.

Example 3h. A plane is missing, and it is presumed that it was equally likely to have gone down in any of 3 possible regions. Let $1 - \beta_i$ denote the probability that the plane will be found upon a search of the ith region when the plane is, in fact, in that region, $i = 1, 2, 3$. (The constants β_i are called overlook probabilities because they represent the probability of overlooking the plane; they are generally attributable to the geographical and environmental conditions of the regions.) What is the conditional probability that the plane is in the ith region, given that a search of region 1 is unsuccessful, $i = 1, 2, 3$?

Solution Let R_i, $i = 1, 2, 3$, be the event that the plane is in region i; and let E be the event that a search of region 1 is unsuccessful. From Bayes' formula we obtain

$$\begin{aligned}
P(R_1|E) &= \frac{P(ER_1)}{P(E)} \\
&= \frac{P(E|R_1)P(R_1)}{\sum_{i=1}^{3} P(E|R_i)P(R_i)} \\
&= \frac{(\beta_1)\frac{1}{3}}{(\beta_1)\frac{1}{3} + (1)\frac{1}{3} + (1)\frac{1}{3}} \\
&= \frac{\beta_1}{\beta_1 + 2}
\end{aligned}$$

For $j = 2, 3$,

$$\begin{aligned}
P(R_j|E) &= \frac{P(E|R_j)P(R_j)}{P(E)} \\
&= \frac{(1)\frac{1}{3}}{(\beta_1)\frac{1}{3} + \frac{1}{3} + \frac{1}{3}} \\
&= \frac{1}{\beta_1 + 2} \qquad j = 2, 3
\end{aligned}$$

It should be noted that the updated (that is, the conditional) probability that the plane is in region j, given the information that a search of region 1 did not find it, is greater than the initial probability that it was in region j when $j \neq 1$ and is less than the initial probability when $j = 1$; which is certainly intuitive, since not finding it when searching region 1 would seem to decrease its chance of being in that region and increase its chance of being elsewhere. Also the conditional probability that the plane is in region 1, given an unsuccessful search of that region, is an increasing function of the overlook probability β_1, which is also intuitive since the larger β_1 is, the more it is reasonable to attribute the unsuccessful search to "bad luck" as opposed to the plane not being there. Similarly, $P(R_j|E)$, $j \neq 1$ is a decreasing function of β_1. ▮

The next example has often been used by unscrupulous probability students to win money from their less enlightened friends.

Example 3i. Suppose that we have 3 cards identical in form except that both sides of the first card are colored red, both sides of the second card are colored black, and one side of the third card is colored red and the other side black. The 3 cards are mixed up in a hat, and 1 card is randomly selected and put down on the ground. If the upper side of the chosen card is colored red, what is the probability that the other side is colored black?

Solution Let RR, BB, and RB denote, respectively, the events that the chosen card is all red, all black, or the red–black card. Letting R be the event that the upturned side of the chosen card is red, we have that the desired probability is obtained by

$$\begin{aligned}
P(RB|R) &= \frac{P(RB \cap R)}{P(R)} \\
&= \frac{P(R|RB)P(RB)}{P(R|RR)P(RR) + P(R|RB)P(RB) + P(R|BB)P(BB)} \\
&= \frac{(\frac{1}{2})(\frac{1}{3})}{(1)(\frac{1}{3}) + (\frac{1}{2})(\frac{1}{3}) + 0(\frac{1}{3})} = \frac{1}{3}
\end{aligned}$$

Hence the answer is $\frac{1}{3}$. Some students guess $\frac{1}{2}$ as the answer by incorrectly reasoning that given that a red side appears, there are two equally likely possibilities: that the card is the all red card or the red–black card. Their mistake, however, is in assuming that these two possibilities are equally likely. For, if we think of each card as consisting of two distinct sides, then there are 6 equally likely outcomes of the experiment—namely, R_1, R_2, B_1, B_2, R_3, B_3—where the outcome is R_1 if the first side of the all red card is turned face up, R_2 if the second side of the all red card is turned face up, R_3 if the red side of the red–black card is turned face up, and so on. Since the other side of the upturned red side will be black only if the outcome is R_3, we see that the desired probability is the conditional probability of R_3 given that either R_1 or R_2 or R_3 occurred, which obviously equals $\frac{1}{3}$. ▮

Example 3j. A new couple, known to have two children, has just moved into town. Suppose that the mother is encountered walking with one of her children. If this child is a girl, what is the probability that both children are girls?

Solution Let us start by defining the following events:

G_1: the first (that is, oldest) child is a girl.
G_2: the second child is a girl.
G: the child seen with the mother is a girl.

Also, let B_1, B_2, B denote similar events except that "girl" is replaced by "boy." Now, the desired probability is $P(G_1G_2|G)$, which can be expressed as follows:

$$\begin{aligned}
P(G_1G_2|G) &= \frac{P(G_1G_2G)}{P(G)} \\
&= \frac{P(G_1G_2)}{P(G)}
\end{aligned}$$

Also,

$$\begin{aligned}
P(G) &= P(G|G_1G_2)P(G_1G_2) + P(G|G_1B_2)P(G_1B_2) \\
&\quad + P(G|B_1G_2)P(B_1G_2) + P(G|B_1B_2)P(B_1B_2) \\
&= P(G_1G_2) + P(G|G_1B_2)P(G_1B_2) + P(G|B_1G_2)P(B_1G_2)
\end{aligned}$$

where the final equation used the results $P(G|G_1G_2) = 1$ and $P(G|B_1B_2) = 0$. If we now make the usual assumption that all 4 gender possibilities are equally likely, then we see that

$$P(G_1G_2|G) = \frac{\frac{1}{4}}{\frac{1}{4} + P(G|G_1B_2)/4 + P(G|B_1G_2)/4}$$
$$= \frac{1}{1 + P(G|G_1B_2) + P(G|B_1G_2)}$$

Thus the answer depends on whatever assumptions we want to make about the conditional probabilities that the child seen with the mother is a girl given the event G_1B_2, and that the child seen with the mother is a girl given the event G_2B_1. For instance, if we want to assume that independent of the genders of the children, the child walking with the mother is the elder child with some probability p, then it would follow that

$$P(G|G_1B_2) = p = 1 - P(G|B_1G_2)$$

implying under this scenario that

$$P(G_1G_2|G) = \tfrac{1}{2}$$

On the other hand, if we were to assume that if the children are of different genders, then the mother would choose to walk with the girl with probability q, independent of the birth order of the children, then we would have that

$$P(G|G_1B_2) = P(G|B_1G_2) = q$$

implying that

$$P(G_1G_2|G) = \frac{1}{1 + 2q}$$

For instance, if we took $q = 1$, meaning that the mother would always choose to walk with a daughter, then the conditional probability of two daughters would be $\frac{1}{3}$, which is in accord with Example 2d because seeing the mother with a daughter is now equivalent to the event that there is at least one daughter.

Thus, as stated, the problem is incapable of solution. Indeed, even when the usual assumption about equally likely gender probabilities is made, we still need to make additional assumptions before a solution can be given. This is because the sample space of the experiment consists of vectors of the form s_1, s_2, i, where s_1 is the gender of the older child, s_2 is the gender of the younger child, and i identifies the birth order of the child seen with the mother. As a result, to specify the probabilities of the events of the sample space, it is not enough to make assumptions only about the genders of the children, it is also necessary to assume something about the conditional probabilities as to which child is with the mother given the genders of the children. ∎

Example 3k. At a psychiatric clinic the social workers are so busy that, on the average, only 60 percent of potential new patients that telephone are able to talk immediately with a social worker when they call. The other 40 percent are asked to leave their phone numbers. About 75 percent of the time a social worker is able to return the call on the same day, and the other 25 percent of the time the caller is contacted on the following day. Experience at the clinic indicates that the probability a caller will actually visit the clinic for consultation is .8 if the caller was immediately able to speak to a social worker, whereas it is .6 and .4, respectively, if the patient's call was returned the same day or the following day.

(a) What percentage of people that telephone visit the clinic for consultation?
(b) What percentage of patients that visit the clinic did not have to have their telephone calls returned?

Solution Define the events V, I, S, F by

V: caller visits the clinic for consultation.
I: caller immediately speaks to a social worker.
S: caller is contacted later on the same day.
F: caller is contacted on the following day.

Then

$$\begin{aligned}P(V) &= P(V|I)P(I) + P(V|S)P(S) + P(V|F)P(F)\\ &= (.8)(.6) + (.6)(.4)(.75) + (.4)(.4)(.25)\\ &= .70\end{aligned}$$

where we have used the facts that $P(S) = (.4)(.75)$ and $P(F) = (.4)(.25)$. Hence part (a) is answered. To answer part (b), we note that

$$\begin{aligned}P(I|V) &= \frac{P(V|I)P(I)}{P(V)}\\ &= \frac{(.8)(.6)}{.7}\\ &\approx .686\end{aligned}$$

Hence approximately 69 percent of the patients that visit the clinic had their phone call immediately answered by a social worker. ∎

3.4 INDEPENDENT EVENTS

The previous examples of this chapter show that $P(E|F)$, the conditional probability of E given F, is not generally equal to $P(E)$, the unconditional probability of E. In other words, knowing that F has occurred generally changes the chances of E's occurrence. In the special cases where $P(E|F)$ does in fact equal $P(E)$, we say that E is independent of F. That is, E is independent of F if knowledge that F has occurred does not change the probability that E occurs.

Since $P(E|F) = P(EF)/P(F)$, we see that E is independent of F if

$$P(EF) = P(E)P(F) \tag{4.1}$$

As Equation (4.1) is symmetric in E and F, it shows that whenever E is independent of F, F is also independent of E. We thus have the following definition.

Definition

Two events E and F are said to be *independent* if Equation (4.1) holds. Two events E and F that are not independent are said to be *dependent*.

Example 4a. A card is selected at random from an ordinary deck of 52 playing cards. If E is the event that the selected card is an ace and F is the event that it is a spade, then E and F are independent. This follows because $P(EF) = \frac{1}{52}$, whereas $P(E) = \frac{4}{52}$ and $P(F) = \frac{13}{52}$. ∎

Example 4b. Two coins are flipped, and all 4 outcomes are assumed to be equally likely. If E is the event that the first coin lands heads and F the event that the second lands tails, then E and F are independent, since $P(EF) = P(\{(H, T)\}) = \frac{1}{4}$; whereas $P(E) = P(\{(H, H), (H, T)\}) = \frac{1}{2}$ and $P(F) = P(\{(H, T), (T, T)\}) = \frac{1}{2}$. ∎

Example 4c. Suppose that we toss 2 fair dice. Let E_1 denote the event that the sum of the dice is 6 and F denote the event that the first die equals 4. Then

$$P(E_1F) = P(\{(4, 2)\}) = \tfrac{1}{36}$$

whereas

$$P(E_1)P(F) = \left(\tfrac{5}{36}\right)\left(\tfrac{1}{6}\right) = \tfrac{5}{216}$$

Hence E_1 and F are not independent. Intuitively, the reason for this is clear because if we are interested in the possibility of throwing a 6 (with 2 dice) we shall be quite happy if the first die lands 4 (or any of the numbers 1, 2, 3, 4, 5), for then we shall still have a possibility of getting a total of 6. On the other hand, if the first die landed 6, we would be unhappy because we would no longer have a chance of getting a total of 6. In other words, our chance of getting a total of six depends on the outcome of the first die; hence E_1 and F cannot be independent.

Now, suppose that we let E_2 be the event that the sum of the dice equals 7. Is E_2 independent of F? The answer is yes, since

$$P(E_2F) = P(\{(4, 3)\}) = \tfrac{1}{36}$$

whereas

$$P(E_2)P(F) = \left(\tfrac{1}{6}\right)\left(\tfrac{1}{6}\right) = \left(\tfrac{1}{36}\right)$$

We leave it for the reader to present the intuitive argument why the event that the sum of the dice equals seven is independent of the outcome on the first die. ∎

Example 4d. If we let E denote the event that the next president is a Republican and F the event that there will be a major earthquake within the next year, then most people would probably be willing to assume that E and F are independent. However, there would probably be some controversy over whether it is reasonable to assume that E is independent of G, where G is the event that there will be a major war within two years after the election. ∎

We now show that if E is independent of F, then E is also independent of F^c.

Proposition 4.1

If E and F are independent, then so are E and F^c.

Proof: Assume that E and F are independent. Since $E = EF \cup EF^c$, and EF and EF^c are obviously mutually exclusive, we have that

$$\begin{aligned} P(E) &= P(EF) + P(EF^c) \\ &= P(E)P(F) + P(EF^c) \end{aligned}$$

or equivalently,

$$\begin{aligned} P(EF^c) &= P(E)[1 - P(F)] \\ &= P(E)P(F^c) \end{aligned}$$

and the result is proved. ∎

Thus, if E is independent of F, then the probability of E's occurrence is unchanged by information as to whether or not F has occurred.

Suppose now that E is independent of F and is also independent of G. Is E then necessarily independent of FG? The answer, somewhat surprisingly, is no. Consider the following example.

Example 4e. Two fair dice are thrown. Let E denote the event that the sum of the dice is 7. Let F denote the event that the first die equals 4 and let G be the event that the second die equals 3. From Example 4c we know that E is independent of F, and the same reasoning as applied there shows that E is also independent of G; but clearly E is not independent of FG [since $P(E|FG) = 1$]. ∎

It would appear to follow from Example 4e that an appropriate definition of the independence of three events E, F, and G would have to go further than merely assuming that all of the $\binom{3}{2}$ pairs of events are independent. We are thus led to the following definition.

Definition

The three events E, F, and G are said to be independent if

$$\begin{aligned}P(EFG) &= P(E)P(F)P(G)\\ P(EF) &= P(E)P(F)\\ P(EG) &= P(E)P(G)\\ P(FG) &= P(F)P(G)\end{aligned}$$

It should be noted that if E, F, and G are independent, then E will be independent of any event formed from F and G. For instance, E is independent of $F \cup G$, since

$$\begin{aligned}P[E(F \cup G)] &= P(EF \cup EG)\\ &= P(EF) + P(EG) - P(EFG)\\ &= P(E)P(F) + P(E)P(G) - P(E)P(FG)\\ &= P(E)[P(F) + P(G) - P(FG)]\\ &= P(E)P(F \cup G)\end{aligned}$$

Of course, we may also extend the definition of independence to more than three events. The events $E_1, E_2, \ldots, E_n$ are said to be independent if, for every subset $E_{1'}, E_{2'}, \ldots, E_{r'}$, $r \leq n$, of these events

$$P(E_{1'}E_{2'} \cdots E_{r'}) = P(E_{1'})P(E_{2'}) \cdots P(E_{r'})$$

Finally, we define an infinite set of events to be independent if every finite subset of these events is independent.

It is sometimes the case that the probability experiment under consideration consists of performing a sequence of subexperiments. For instance, if the experiment consists of continually tossing a coin, we may think of each toss as being a subexperiment. In many cases it is reasonable to assume that the outcomes of any group of the subexperiments have no effect on the probabilities of the outcomes of the other subexperiments. If such is the case, we say that the subexperiments are independent. More formally, we say that the subexperiments are independent if $E_1, E_2, \ldots, E_n, \ldots$ is necessarily an independent sequence of events whenever E_i is an event whose occurrence is completely determined by the outcome of the ith subexperiment.

If each subexperiment is identical—that is, if each subexperiment has the same (sub) sample space and the same probability function on its events—then the subexperiments are called *trials*.

Example 4f. An infinite sequence of independent trials is to be performed. Each trial results in a success with probability p and a failure with probability $1 - p$. What is the probability that

(a) at least 1 success occurs in the first n trials;
(b) exactly k successes occur in the first n trials;
*(c) all trials result in successes?

Solution In order to determine the probability of at least 1 success in the first n trials, it is easiest to compute first the probability of the complementary event, that of no successes in the first n trials. If we let E_i denote the event of a failure on the ith trial, then the probability of no successes is, by independence,

$$P(E_1E_2 \cdots E_n) = P(E_1)P(E_2) \cdots P(E_n) = (1 - p)^n$$

Hence the answer to part (a) is $1 - (1 - p)^n$.

To compute part (b), consider any particular sequence of the first n outcomes containing k successes and $n - k$ failures. Each one of these sequences will, by the assumed independence of trials, occur with probability $p^k(1 - p)^{n-k}$. As there are $\binom{n}{k}$ such sequences (there are $n!/k!\,(n - k)!$ permutations of k successes and $n - k$ failures), the desired probability in part (b) is

$$P\{\text{exactly } k \text{ successes}\} = \binom{n}{k}p^k(1 - p)^{n-k}$$

To answer part (c), we note by part (a) that the probability of the first n trials all resulting in successes is given by

$$P(E_1^cE_2^c \cdots E_n^c) = p^n$$

Hence, using the continuity property of probabilities (Section 2.6), we have that the desired probability $P\left(\bigcap_1^\infty E_i^c\right)$ is given by

$$\begin{aligned}
P\left(\bigcap_{i=1}^{\infty} E_i^c\right) &= P\left(\lim_{n\to\infty} \bigcap_{i=1}^{n} E_i^c\right) \\
&= \lim_{n\to\infty} P\left(\bigcap_{i=1}^{n} E_i^c\right) \\
&= \lim_{n} p^n = \begin{cases} 0 & \text{if } p < 1 \\ 1 & \text{if } p = 1 \end{cases}
\end{aligned}$$

▮

Example 4g. A system composed of n separate components is said to be a parallel system if it functions when at least one of the components functions (see Figure 3.2). For such a system, if component i, independent of other components, functions with probability p_i, $i = 1, \ldots, n$, what is the probability that the system functions?

Solution Let A_i denote the event that component i functions. Then

$$\begin{aligned}
P\{\text{system functions}\} &= 1 - P\{\text{system does not function}\} \\
&= 1 - P\{\text{all components do not function}\} \\
&= 1 - P\left(\bigcap_{i} A_i^c\right) \\
&= 1 - \prod_{i=1}^{n} (1 - p_i) \qquad \text{by independence}
\end{aligned}$$

▮

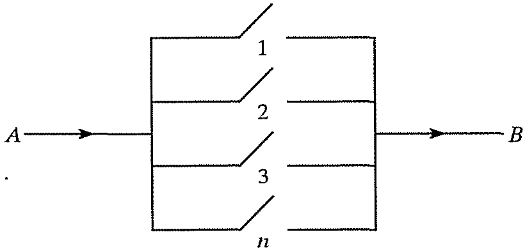


Figure 3.2 Parallel system: functions if current flows from A to B.

Example 4h. Independent trials, consisting of rolling a pair of fair dice, are performed. What is the probability that an outcome of 5 appears before an outcome of 7 when the outcome of a roll is the sum of the dice?

Solution If we let E_n denote the event that no 5 or 7 appears on the first $n - 1$ trials and a 5 appears on the nth trial, then the desired probability is

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

Now, since $P\{5 \text{ on any trial}\} = \frac{4}{36}$ and $P\{7 \text{ on any trial}\} = \frac{6}{36}$, we obtain, by the independence of trials

$$P(E_n) = \left(1 - \tfrac{10}{36}\right)^{n-1} \tfrac{4}{36}$$

and thus

$$\begin{aligned} P\left(\bigcup_{n=1}^{\infty} E_n\right) &= \frac{1}{9} \sum_{n=1}^{\infty} \left(\frac{13}{18}\right)^{n-1} \\ &= \frac{1}{9} \frac{1}{1 - \frac{13}{18}} \\ &= \frac{2}{5} \end{aligned}$$

This result may also have been obtained by using conditional probabilities. If we let E be the event that a 5 occurs before a 7, then we can obtain the desired probability, $P(E)$, by conditioning on the outcome of the first trial, as follows: Let F be the event that the first trial results in a 5; let G be the event that it results in a 7; and let H be the event that the first trial results in neither a 5 nor a 7. Conditioning on which one of these events occurs gives

$$P(E) = P(E|F)P(F) + P(E|G)P(G) + P(E|H)P(H)$$

However,

$$\begin{aligned} P(E|F) &= 1 \\ P(E|G) &= 0 \\ P(E|H) &= P(E) \end{aligned}$$

The first two equalities are obvious. The third follows because if the first outcome results in neither a 5 nor a 7, then at that point the situation is exactly as when the problem first started; namely, the experimenter will continually roll a pair of fair dice until either a 5 or 7 appears. Furthermore, the trials are independent; therefore, the outcome of the first trial will have no effect on subsequent rolls of dice. Since $P(F) = \frac{4}{36}$, $P(G) = \frac{6}{36}$, and $P(H) = \frac{26}{36}$, we see that

$$P(E) = \tfrac{1}{9} + P(E)\tfrac{13}{18}$$

or

$$P(E) = \tfrac{2}{5}$$

The reader should note that the answer is quite intuitive. That is, since a 5 occurs on any roll with probability $\frac{4}{36}$ and a 7 with probability $\frac{6}{36}$, it seems intuitive that the odds that a 5 appears before a 7 should be 6 to 4 against. The probability should be $\frac{4}{10}$, as indeed it is.

The same argument shows that if E and F are mutually exclusive events of an experiment, then, when independent trials of this experiment are performed, the event E will occur before the event F with probability

$$\frac{P(E)}{P(E) + P(F)}$$ ▮

The next example presents a problem that occupies an honored place in the history of probability theory. This is the famous *problem of the points*. In general terms, the problem is this: Two players put up stakes and play some game, with the stakes to go to the winner of the game. An interruption requires them to stop before either has won, and when each has some sort of a "partial score." How should the stakes be divided?

This problem was posed to the French mathematician Pascal in 1654 by the Chevalier de Méré, who was a professional gambler at that time. In attacking the problem, Pascal introduced the important idea that the proportion of the prize deserved by the competitors should depend on their respective probabilities of winning if the game were to be continued at that point. Pascal worked out some special cases, and, more important, initiated a correspondence with the famous Frenchman Fermat, who had a great reputation as a mathematician. The resulting exchange of letters led not only to a complete solution to the problem of the points, but also laid the framework for the solution to many other problems connected with games of chance. This celebrated correspondence, dated by some as the birth date of probability theory, was also important in stimulating interest in probability among the mathematicians in Europe, for Pascal and Fermat were both recognized as being among the foremost mathematicians of the time. For instance, within a short time of their correspondence, the young Dutch mathematician Huygens came to Paris to discuss these problems and solutions; and interest and activity in this new field grew rapidly.