### Key Concepts

<table>
<thead>
<tr>
<th>Three definitions of probability</th>
<th>Conditional probability, mutually exclusive events, independent events</th>
</tr>
</thead>
<tbody>
<tr>
<td>The probability of either of two events, the joint probability of two events</td>
<td>Bayes’ theorem</td>
</tr>
<tr>
<td></td>
<td>Likelihood ratio</td>
</tr>
</tbody>
</table>
SYMBOLES AND ABBREVIATIONS

\[ P(A) \] probability of the event \( A \)
\[ P(A \text{ or } B) \] probability of either event \( A \) or event \( B \)
\[ P(A \text{ and } B) \] joint probability of the events \( A \) and \( B \)
\[ P(A|B) \] conditional probability of event \( A \) given event \( B \)

DEFINITION OF PROBABILITY

Although the meaning of a statement such as ‘The probability of rain today is 50%’ may seem fairly obvious, it is not easy to give an exact definition of the term probability. In fact, three different definitions have been proposed, each focusing on a different aspect of the concept. Since probability plays such a fundamental role in genetics and in statistics in general, we start by reviewing all three of these definitions before stating the mathematical laws that govern its manipulation. The mathematicians who originally studied probability were motivated by gambling and so used games of chance (e.g. cards and dice) in their studies. It will be convenient for us to use similar examples initially.

The classical definition of probability can be stated as follows: Given a set of equally likely possible outcomes, the probability of the event \( A \), which for brevity we write \( P(A) \), is the number of outcomes that are ‘favorable to’ \( A \) divided by the total number of possible outcomes:

\[
P(A) = \frac{\text{number of outcomes favorable to } A}{\text{total number of possible outcomes}}.
\]

This definition will become clearer when we illustrate it with some examples. Suppose we pick a single card at random from a well-shuffled regular deck of 52 cards comprising four suits (clubs, diamonds, hearts and spades) with 13 cards in each suit. Note that when we say a card is drawn at random from a well-shuffled deck of
52 cards, we imply that each card is equally likely to be selected so that each of the 52 cards has probability $1/52$ of being the one selected. What is the probability that the card is an ace? We let $A$ be the event ‘the card is an ace’. Each card represents a possible outcome, and so the total number of possible outcomes is 52; the number of outcomes that are ‘favorable to’ $A$ is 4, since there are four aces in the deck. Therefore the probability is

$$P(A) = \frac{4}{52} = \frac{1}{13}.$$ 

As another example, what is the probability of obtaining two sixes when two normal six-sided dice are rolled? In this example, let $A$ be the event ‘obtaining two sixes’. Each of the two dice can come up one, two, three, four, five, or six, and so the total number of possible outcomes is $36$ ($6 \times 6$). Only one of these outcomes (six and six) is ‘favorable to’ $A$, and so the probability is $P(A) = 1/36$.

This definition of probability is precise and appears to make sense. Unfortunately, however, it contains a logical flaw. Note that the definition includes the words ‘equally likely’, which is another way of saying ‘equally probable’. In other words, probability has been defined in terms of probability! Despite this difficulty, our intuition tells us that when a card is taken at random from a well-shuffled deck, or when normal dice are rolled, there is physical justification for the notion that all possible outcomes are ‘equally likely’. When there is no such physical justification (as we shall see in a later example), the classical definition can lead us astray if we are not careful.

The frequency definition of probability supposes that we can perform many, many replications or trials of the same experiment. As the number of trials tends to infinity (i.e. to any very large number, represented mathematically by the symbol $\infty$), the proportion of trials in which the event $A$ occurs tends to a fixed limit. We then define the probability of the event $A$ as this limiting proportion. Thus, to answer the question, ‘What is the probability that the card is an ace?’, We suppose that many trials are performed, in each of which a card is drawn at random from a well-shuffled deck of 52 cards. After each trial we record in what proportion of the cards drawn so far an ace has been drawn, and we find that, as the number of trials increases indefinitely, this proportion tends to the limiting value of $1/13$. This concept is illustrated in Figure 4.1, which represents a set of trials in which an ace was drawn at the 10th, 20th, 40th, 54th, 66th, and 80th trials.

Since one can never actually perform an infinite number of trials, this definition of probability is mathematically unsatisfying. Nevertheless, it is the best way of interpreting probability in practical situations. The statement ‘There is a 50% probability of rain today’ can be interpreted to mean that on just half of many days on which such a statement is made, it will rain. Provided this is true, the probability statement is valid.
Figure 4.1 Example of probability defined as the limiting proportion of trials, as the number of trials tends of infinity, in which a particular event occurs.

In genetic counseling, however, quoting valid probabilities may not be sufficient. When a pregnant woman is counseled concerning the probability that her child will have a particular disease, the probability that is quoted must be both valid and relevant. Suppose, for example, a 25-year-old woman has a child with Down syndrome (trisomy 21) and, on becoming pregnant again, seeks counseling. Her physician checks published tables and finds that the probability of a 25-year-old woman having a Down syndrome child is about 1 in 400. If she is counseled that the probability of her second child having Down syndrome is only about 1 in 400, she will have been quoted a valid probability; only 1 in 400 such women coming for counseling will bear a child with Down syndrome. But this probability will not be relevant if the first child has Down syndrome because of a translocation in the mother, that is, because the mother has a long arm of chromosome 21 attached to another of her chromosomes. If this is the case, the risk to the unborn child is about 1 in 6. A physician who failed to recommend the additional testing required to arrive at such a relevant probability for a specific patient could face embarrassment, perhaps even a malpractice suit, though the quoted probability, on the basis of all the information available, was perfectly valid.

The mathematical (axiomatic) definition of probability avoids the disadvantages of the two other definitions and is the definition used by mathematicians. A simplified version of this definition, which, although incomplete, retains its main features, is as follows. A set of probabilities is any set of numbers for which: each number is greater than or equal to zero; and the sum of the numbers is unity (one). This definition, unlike the other two, gives no feeling for the practical meaning.
of probability. It does, however, describe the essential characteristics of probability: there is a set of possible outcomes, each associated with a positive probability of occurring, and at least one of these outcomes must occur. We now turn to some fundamental laws of probability, which do not depend on which definition is taken.

**THE PROBABILITY OF EITHER OF TWO EVENTS: \( A \) OR \( B \)**

If \( A \) and \( B \) are two events, what is the probability of either \( A \) or \( B \) occurring? That is, what is the probability that \( A \) occurs but \( B \) does not, \( B \) occurs but \( A \) does not, or both \( A \) and \( B \) occur? Let us consider a simple example. A regular deck of cards is shuffled well and a card is drawn. What is the probability that it is either an ace or a king? The deck contains eight cards that are either aces or kings, and so the answer is \( \frac{8}{52} = \frac{2}{13} \). Now notice that the two events ‘ace’ and ‘king’ are mutually exclusive, in that when we draw a card from the deck, it cannot be both an ace and a king. If \( A \) and \( B \) are mutually exclusive events, then

\[
P(\text{A or B}) = P(A) + P(B).
\]

Thus, in our example, we have

\[
P(\text{ace or king}) = P(\text{ace}) + P(\text{king}) = \frac{1}{13} + \frac{1}{13} = \frac{2}{13}.
\]

Now suppose the question had been: ‘What is the probability that the card is an ace or a heart?’ In this case the events ‘ace’ and ‘heart’ are not mutually exclusive, because the same card could be both an ace and a heart; therefore we cannot use the formula given above. How many cards in the deck are either aces or hearts? There are 4 aces and 13 hearts, but a total of only 16 cards that are either an ace or a heart: the ace of clubs, the ace of diamonds, the ace of spades, and the 13 hearts. Notice that if we carelessly write ‘4 aces + 13 hearts = 17 cards’, the ace of hearts has been counted twice – once as an ace and once as a heart. In other words, the number of cards that are either aces or hearts is the number of aces, plus the number of hearts, minus the number of cards that are both aces and hearts (one, in
this example). Analogously, dividing each of these numbers by 52 (the total number of cards in the deck), we have

\[ P(\text{ace or heart}) = P(\text{ace}) + P(\text{heart}) - P(\text{ace and heart}) \]

\[ = \frac{4}{52} + \frac{13}{52} - \frac{1}{52} = \frac{16}{52} = \frac{4}{13}. \]

The general rule for any two events \( A \) and \( B \) is

\[ P(\text{A or B}) = P(\text{A}) + P(\text{B}) - P(\text{A and B}). \]

This rule is general, by which we mean it is always true. (A layman uses the word ‘generally’ to mean ‘usually.’ In mathematics a general result is one that is always true, and the word ‘generally’ means ‘always.’)

If the event \( A \) is ‘the card is a king’ and the event \( B \) is ‘the card is an ace’, we have

\[ P(\text{A or B}) = P(\text{A}) + P(\text{B}) - P(\text{A and B}) \]

\[ = \frac{1}{13} + \frac{1}{13} - 0 = \frac{2}{13} \]

as before. In this example, \( P(\text{A and B}) = 0 \) because a card cannot be both a king and an ace. In the special case in which \( A \) and \( B \) are mutually exclusive events, \( P(\text{A and B}) = 0. \)

**The Joint Probability of Two Events: \( A \) and \( B \)**

We have just seen that the probability that both events \( A \) and \( B \) occur is written \( P(\text{A and B}) \); this is also sometimes abbreviated to \( P(AB) \) or \( P(A,B) \). It is often called the joint probability of \( A \) and \( B \). If \( A \) and \( B \) are mutually exclusive (i.e. they cannot both occur – if a single card is drawn, for example, it cannot be both an ace and a king), then their joint probability, \( P(\text{A and B}) \), is zero. What can we say about \( P(\text{A and B}) \) in general? One answer to this question is implicit in the general formula for \( P(\text{A or B}) \) just given. Rearranging this formula we find

\[ P(\text{A and B}) = P(\text{A}) + P(\text{B}) - P(\text{A or B}). \]

A more useful expression, however, uses the notion of conditional probability: the probability of an event occurring given that another event has already occurred. We write the conditional probability of \( B \) occurring given that \( A \) has occurred as
Read the vertical line as ‘given’, so that \( P(B|A) \) is the ‘probability of \( B \) given \( A \)’ (i.e. given that \( A \) has already occurred). Sensitivity and specificity are examples of conditional probabilities. We have

\[
\text{Sensitivity} = P(\text{test positive}|\text{disease present}), \\
\text{Specificity} = P(\text{test negative}|\text{disease absent}).
\]

Using this concept of conditional probability, we have the following general rule for the joint probability of \( A \) and \( B \):

\[
P(A \text{ and } B) = P(A)P(B|A).
\]

Since it is arbitrary which event we call \( A \) and which \( B \), note that we could equally well have written

\[
P(A \text{ and } B) = P(B)P(A|B).
\]

Using this formula, what is the probability, on drawing a single card from a well-shuffled deck, that it is both an ace and a heart (i.e. that it is the ace of hearts)? Let \( A \) be the event that the card is an ace and \( B \) be the event that it is a heart. Then, from the formula,

\[
P(\text{ace and heart}) = P(\text{ace})P(\text{heart}|\text{ace})
\]

\[
= \frac{1}{13} \times \frac{1}{4} = \frac{1}{52}.
\]

\( P(\text{heart}|\text{ace}) = 1/4 \), because the ace that we have picked can be any of the four suits, only one of which is hearts. Similarly, we could have found

\[
P(\text{ace and heart}) = P(\text{heart})P(\text{ace}|\text{heart})
\]

\[
= \frac{1}{4} \times \frac{1}{13} = \frac{1}{52}.
\]

Now notice that in this particular example we have

\[
P(\text{heart}|\text{ace}) = P(\text{heart}) = \frac{1}{4}
\]

and

\[
P(\text{ace}|\text{heart}) = P(\text{ace}) = \frac{1}{13}.
\]
In other words, the probability of picking a heart is the same (1/4) whether or not an ace has been picked, and the probability of picking an ace is the same (1/13) whether or not a heart has been picked. These two events are therefore said to be independent.

Two events \( A \) and \( B \) are independent if \( P(A) = P(A|B) \), or if \( P(B) = P(B|A) \). From the general formula for \( P(A \text{ and } B) \), it follows that if two events \( A \) and \( B \) are independent, then

\[
P(A \text{ and } B) = P(A)P(B).
\]

Conversely, two events \( A \) and \( B \) are independent if we know that

\[
P(A \text{ and } B) = P(A)P(B).
\]

It is often intuitively obvious when two events are independent. Suppose we have two regular decks of cards and randomly draw one card from each. What is the probability that the card from the first deck is a king and the card from the second deck is an ace? The two draws are clearly independent, because the outcome of the first draw cannot in any way affect the outcome of the second draw. The probability is thus

\[
\frac{1}{13} \times \frac{1}{13} = \frac{1}{169}.
\]

But suppose we have only one deck of cards, from which we draw two cards consecutively (where after the first card is drawn, it is not put back in the deck before the other card is drawn so that the second card is drawn from a deck containing only 51 cards). Now what is the probability that the first is a king and the second is an ace? Using the general formula for the joint probability of two events \( A \) and \( B \), we have

\[
P(\text{lst is king and 2nd is ace}) = P(\text{lst is king})P(\text{2nd is ace}|\text{1st is king})
\]

\[
= \frac{4}{52} \times \frac{4}{51} = \frac{4}{663}.
\]

In this case the two draws are not independent. The probability that the second card is an ace depends on what the first card is (if the first card is an ace, for example, the probability that the second card is an ace becomes 3/51).
EXAMPLES OF INDEPENDENCE, NONINDEPENDENCE AND GENETIC COUNSELING

It is a common mistake to assume two events are independent when they are not. Suppose two diseases occur in a population and it is known that it is impossible for a person to have both diseases. There is a strong temptation to consider two such diseases to be occurring independently in the population, whereas in fact this is impossible. Can you see why this is so? [Hint: Let the occurrence of one disease in a particular individual be the event $A$, and the occurrence of the other disease be event $B$. What do you know about the joint probability of $A$ and $B$ if (1) they are independent, and (2) they are mutually exclusive? Can $P(A \text{ and } B)$ be equal to both $P(A)P(B)$ and zero if $P(A)$ and $P(B)$ are both nonzero (we are told both diseases actually occur in the population)?] On the other hand, it is sometimes difficult for an individual to believe that truly independent events occur in the manner in which they do occur. The mother of a child with a genetic anomaly may be properly counseled that she has a 25% probability of having a similarly affected child at each conception and that all conceptions are independent. But she will be apt to disbelieve the truth of such a statement if (as will happen to one quarter of mothers in this predicament, assuming the counseling is valid and she has another child) her very next child is affected.

NONINDEPENDENCE DUE TO MULTIPLE ALLELISM

Among the Caucasian population, 44% have red blood cells that are agglutinated by an antiserum denoted anti-A, and 56% do not. Similarly, the red blood cells of 14% of the population are agglutinated by another antiserum denoted anti-B, and those of 86% are not. If these two traits are distributed independently in the population, what proportion would be expected to have red cells that are agglutinated by both anti-A and anti-B? Let $A+$ be the event that a person’s red cells are agglutinated by anti-A, and $B+$ the event that they are agglutinated by anti-B. Thus, $P(A+) = 0.44$ and $P(B+) = 0.14$. If these two events are independent, we should expect

$$P(A+ \text{ and } B+) = P(A+)P(B+) = 0.44 \times 0.14 = 0.06.$$ 

In reality, less than 4% of Caucasians fall into this category (i.e. have an AB blood type); therefore, the two traits are not distributed independently in the population. This finding was the first line of evidence used to argue that these two traits are due to multiple alleles at one locus (the ABO locus), rather than to segregation at two separate loci.
NONINDEPENDENCE DUE TO LINKAGE DISEQUILIBRIUM

The white cells of 31% of the Caucasian population react positively with HLA anti-A1, and those of 21% react positively with HLA anti-B8. If these two traits are independent, we expect the proportion of the population whose cells would react positively to both antisera to be \(0.31 \times 0.21 = 0.065\) (i.e. 6.5%). In fact, we find that 17% of the people in the population are both A1 positive and B8 positive. In this case family studies have shown that two loci (A and B), very close together on chromosome 6, are involved. Nonindependence between two loci that are close together is termed *linkage disequilibrium*.

These two examples illustrate the fact that more than one biological phenomenon can lead to a lack of independence. In many cases in the literature, nonindependence is established, and then, on the basis of that evidence alone, a particular biological mechanism is incorrectly inferred. In fact, as we shall see in Chapter 9, nonindependence (or association) can be due to nothing more than the population comprising two subpopulations.

CONDITIONAL PROBABILITY IN GENETIC COUNSELING

We shall now consider a very simple genetic example that will help you learn how to manipulate conditional probabilities. First, recall the formula for the joint probabilities of events \(A\) and \(B\), which can be written (backward) as

\[
P(A)P(B|A) = P(A\text{ and } B).
\]

Now divide both sides by \(P(A)\), which may be done provided \(P(A)\) is not zero, and we find that

\[
P(B|A) = \frac{P(A\text{ and } B)}{P(A)}.
\]

This gives us a formula for calculating the conditional probability of \(B\) given \(A\), denoted \(P(B|A)\), if we know \(P(A\text{ and } B)\) and \(P(A)\). Similarly, if we know \(P(A\text{ and } B)\) and \(P(B)\), we can find

\[
P(A|B) = \frac{P(A\text{ and } B)}{P(B)} \text{ assuming } P(B) \text{ is not zero}.
\]
Now consider, as an example, hemophilia, caused by a defect in the blood-clotting system. This is a rare X-linked recessive disease (i.e., an allele at a locus on the X-chromosome is recessive with respect to the disease, with the result that a female requires two such alleles to have the disease but a male only one – because a male has only one X chromosome). Suppose a carrier mother, who has one disease allele but not the disease, marries an unaffected man. She will transmit the disease to half her sons and to none of her daughters (half her daughters will be carriers, but none will be affected with the disease). What is the probability that she will bear an affected child? The child must be either a son or a daughter, and these are mutually exclusive events. Therefore,

\[ P(\text{affected child}) = P(\text{affected son}) + P(\text{affected daughter}) \]

\[ = P(\text{son and affected}) + P(\text{daughter and affected}) \]

\[ = P(\text{son})P(\text{affected|son}) + P(\text{daughter})P(\text{affected|daughter}). \]

Assume a 1:1 sex ratio (i.e., \( P(\text{son}) = P(\text{daughter}) = \frac{1}{2} \)), and use the fact that \( P(\text{affected|son}) = \frac{1}{2} \). Furthermore, in the absence of mutation (which has such a small probability that we shall ignore it), \( P(\text{affected|daughter}) = 0 \). We therefore have

\[ P(\text{affected child}) = \frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times 0 = \frac{1}{4}. \]

Now suppose an amniocentesis is performed and thus the gender of the fetus is determined. If the child is a daughter, the probability of her being affected is virtually zero. If, on the other hand, the child is a son, the probability of his being affected is one half. Note that we can derive this probability by using the formula for conditional probability:

\[ P(\text{affected|son}) = \frac{P(\text{son and affected})}{P(\text{son})} \]

\[ = \frac{1/4}{1/2} = \frac{1}{2}. \]

Of course in this case you do not need to use the formula to obtain the correct answer, but you should nevertheless be sure to understand the details of this example. Although a very simple example, it illustrates how one proceeds in more complicated examples. Notice that knowledge of the gender of the child changes the probability of having an affected child from 1/4 (before the gender was known) either to 1/2 (in the case of a son) or to 0 (in the case of a daughter). Conditional probabilities
are used extensively in genetic counseling. We shall now discuss the use of a very general theorem that gives us a mechanism for adjusting the probabilities of events as more information becomes available.

**BAYES’ THEOREM**

The Englishman Thomas Bayes wrote an essay on probability that he did not publish – perhaps because he recognized the flaw in assuming, as he did in his essay, that all possible outcomes are equally likely (this explanation of why he did not publish is disputed). The essay was nevertheless published in 1763, after his death, by a friend. What is now called Bayes’ theorem does not contain this flaw. The theorem gives us a method of calculating new probabilities to take account of new information. Suppose that 20% of a particular population has a certain disease, \( D \). For example, the disease might be hypertension, defined as having an average diastolic blood pressure of 95 mmHg or greater taken once a day over a period of 5 days.

![Figure 4.2](image)

**Figure 4.2** In the whole population, represented by a square whose sides are unity, the probability of having the disease is \( P(D) \), and of not having the disease is \( P(\bar{D}) \), as indicated along the horizontal axis.

In Figure 4.2 we represent the whole population by a square whose sides are unity. The probability that a person has the disease, \( P(D) \), and the probability that a
person does not have the disease, \( P(\overline{D}) = 1 - P(D) \), are indicated along the bottom axis. Thus the areas of the two rectangles are the same as these two probabilities.

Now suppose we have a test that picks up a particular symptom \( S \) associated with the disease. In our example, the test might be to take just one reading of the diastolic blood pressure, and \( S \) might be defined as this one pressure being 95 mmHg or greater. Alternatively, we could say that the test result is positive if this one blood pressure is 95 mmHg or greater, negative otherwise. Before being tested, a random person from the population has a 20% probability of having the disease. How does this probability change if it becomes known that the symptom is present?

**Figure 4.3** Within each subpopulation, \( D \) and \( \overline{D} \), the conditional probability of having a positive test result or symptom, \( S \), is indicated along the vertical axis. The rectangles represent the joint probabilities \( P(D)P(S|D) = P(S,D) \) and \( P(\overline{D})P(S|\overline{D}) = P(S, \overline{D}) \).

Assume that the symptom is present in 90% of all those with the disease but only 10% of all those without the disease, that is, \( P(S|D) = 0.9 \) and \( P(S|\overline{D}) = 0.1 \). In other words, the sensitivity and the specificity of the test are both 0.9. These conditional probabilities are indicated along the vertical axis in Figure 4.3. The gray rectangles represent the joint probabilities that the symptom is present and that the disease is present or not:

\[
P(S \text{ and } D) = P(D)P(S|D) = 0.2 \times 0.9 = 0.18,
\]
\[
P(S \text{ and } \overline{D}) = P(\overline{D})P(S|\overline{D}) = 0.8 \times 0.1 = 0.08.
\]
If we know that the symptom is present, then we know that only the gray areas are relevant, that is, we can write, symbolically,

\[
P(D|S) = \frac{P(S \text{ and } D)}{P(S \text{ and } D) + P(S \text{ and } \overline{D})} = \frac{P(D)P(S|D)}{P(D)P(S|D) + P(\overline{D})P(S|\overline{D})}
\]

\[
= \frac{0.2 \times 0.9}{0.2 \times 0.9 + 0.8 \times 0.1} = 0.69,
\]

which is the positive predictive value of the test. This, in essence, is Bayes’ theorem. We start with a prior probability of the disease, \(P(D)\), which is then converted into a posterior probability, \(P(D|S)\), given the new knowledge that the symptom \(S\) is present.

More generally, we can give the theorem as follows. Let the new information that is available be that the event \(S\) occurred. Now suppose the event \(S\) can occur in anyone of \(k\) distinct, mutually exclusive ways. Call these ways \(D_1, D_2, \ldots, D_k\) (in the above example there were just two ways, the person either had the disease or did not have the disease; in general, there may be \(k\) alternative diagnoses possible). Suppose that with no knowledge about \(S\), these have prior probabilities \(P(D_1), P(D_2), \ldots,\) and \(P(D_k)\), respectively. Then the theorem states that the posterior probability of a particular \(D\), say \(D_j\), conditional on \(S\) having occurred, is

\[
P(D_j|S) = \frac{P(D_j)P(S|D_j)}{P(D_1)P(S|D_1) + P(D_2)P(S|D_2) + \cdots + P(D_k)P(S|D_k)}
\]

\[
= \frac{P(D_j \text{ and } S)}{P(D_1 \text{ and } S) + P(D_2 \text{ and } S) + \cdots + P(D_k \text{ and } S)}.
\]

The theorem can thus be remembered as ‘the joint probability of interest divided by the sum of all the joint probabilities’ (i.e. the posterior probability of a particular \(D\), given that \(S\) has occurred, is equal to the joint probability of \(D\) and \(S\) occurring, divided by the sum of the joint probabilities of each of the \(D\)s and \(S\) occurring). This is illustrated in Figure 4.4.
Figure 4.4  Bayes’ theorem. The probabilities of various diagnoses, $D_1, D_2, \ldots, D_j, \ldots, D_k$, are indicated on the horizontal axis and the conditional probability of a particular symptom, within each diagnostic class $D_j$, is indicated on the vertical axis. Thus each hatched rectangle is the joint probability of the symptom and a diagnostic class.

In order to illustrate how widely applicable Bayes’ theorem is, we shall consider some other examples. First, suppose that a woman knows that her mother carries the allele for hemophilia (because her brother and her maternal grandfather both have the disease). She is pregnant with a male fetus and wants to know the probability that he will be born with hemophilia. There is a $1/2$ probability that she has inherited the hemophilia allele from her mother, and a $1/2$ probability that (given that she inherited the allele) she passes it on to her son. Thus, if this is the only information available, the probability that her son will have hemophilia is $1/4$. This is illustrated in Figure 4.5(a). Now suppose we are told that she already has a son who does not have hemophilia (Figure 4.5(b)). What is now the probability that her next son will be born with hemophilia? Is it the same, is it greater than $1/4$, or is it less than $1/4$? That she has already had a son without hemophilia is new information that has a direct bearing on the situation. In order to see this, consider a more extreme situation: suppose she has already had 10 unaffected sons. This would suggest that she did not inherit the allele from her mother, and if that is the case she could not pass it on to her future sons. If, on the other hand, she already has had a son with hemophilia, she would know without a doubt that she had inherited the allele from her mother and every future son would have a $1/2$ probability of being affected. Thus, the fact that she has had one son who is unaffected decreases the probability that she inherited the hemophilia allele, and hence the probability that her second
son will be affected. We shall now use Bayes’ theorem to calculate the probability that the woman inherited the allele for hemophilia, given that she has a son without hemophilia. We have $k = 2$ and define the following events:

- $S =$ the woman has an unaffected son,
- $D_1 =$ the woman inherited the hemophilia allele,
- $D_2 =$ the woman did not inherit the hemophilia allele.

Before we know that she has an unaffected son, we have the prior probabilities

$$P(D_1) = P(D_2) = \frac{1}{2}.$$ 

Given whether or not she inherited the hemophilia allele, we have the conditional probabilities

$$P(S|D_1) = \frac{1}{2}, \quad P(S|D_2) = 1.$$ 

Therefore, applying Bayes’ theorem, the posterior probability that she inherited the hemophilia allele is

$$P(D_1|S) = \frac{P(D_1) P(S|D_1)}{P(D_1) P(S|D_1) + P(D_2) P(S|D_2)} = \frac{\frac{1}{2} \times \frac{1}{2}}{\frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times 1} = \frac{1}{3}.$$
Thus, given all that we know, the probability that the woman inherited the hemophilia allele is 1/3. Therefore, the probability that her second son is affected, given all that we know, is one half of this, 1/6. Note that $P(S) = 1/6$ is less than the 1/4 probability that would have been appropriate if she had not already had a normal son. In general, the more unaffected sons we know she has, the smaller the probability that her next son will be affected.

Let us take as another example a situation that could arise when there is suspected non-paternity. Suppose a mother has blood type A (and hence genotype AA or AO), and her child has blood type AB (and hence genotype AB). Thus, the child’s A allele came from the mother and the B allele must have come from the father. The mother alleges that a certain man is the father, and his blood is typed. Consider two possible cases:

CASE 1

The alleged father has blood type O, and hence genotype OO. Barring a mutation, he could not have been the father: this is called an exclusion. Where there is an exclusion, we do not need any further analysis.

CASE 2

The alleged father has blood type AB, and therefore genotype AB, which is relatively rare in the population. Thus, not only could the man be the father, but he is also more likely to be the father than a man picked at random (who is much more likely to be O or A, and hence not be the father). In this case we might wish to determine, on the basis of the evidence available, the probability that the alleged father is in fact the true father. We can use Bayes’ theorem for this purpose, provided we are prepared to make certain assumptions:

1. Assume there have been no blood-typing errors and no mutation, and that we have the correct mother. It follows from this assumption that the following two events occurred: (i) the alleged father is AB; and (ii) the child received B from the true father (call this event S).

2. Assume we can specify the different, mutually exclusive ways in which S could have occurred. For the purposes of this example we shall suppose there are just two possible ways in which event S could have occurred (i.e. $k = 2$ again in the theorem):

   (i) The alleged father is the true father ($D_1$). We know from Mendelian genetics that the probability of a child receiving a B allele from an AB father is 0.5, that is, $P(S|D_1) = 0.5$. 
(ii) Or a random man from a specified population is the true father ($D_2$). We shall assume that the frequency of the allele B in this population is 0.06, so that $P(S|D_2) = 0.06$. The probability that a B allele is passed on to a child by a random man from the population is the same as the probability that a random allele in the population at the ABO locus is B (i.e. the population allele frequency of B).

3. Assume we know the prior probabilities of these two possibilities. As explained in the Appendix, these could be estimated from the previous experience of the laboratory doing the blood typing. We shall assume it is known that 65% of alleged fathers whose cases are typed in this laboratory are in fact the true fathers of the children in question, that is, $P(D_1) = 0.65$ and $P(D_2) = 0.35$.

We are now ready to use the formula substituting $P(D_1) = 0.65$, $P(D_2) = 0.35$, $P(S|D_1) = 0.5$ and $P(S|D_2) = 0.06$. Thus we obtain

$$P(D_1|S) = \frac{0.65 \times 0.5}{0.65 \times 0.5 + 0.35 \times 0.06} = \frac{0.325}{0.325 + 0.021} = 0.94.$$ 

A summary of this application of Bayes’ theorem is shown by means of a tree diagram in Figure 4.6.

![Figure 4.6](image-url)  

**Figure 4.6** Calculation of the probability that the father of a child is the alleged father, on the basis of ABO blood types, using Bayes’ theorem.
Thus, whereas the alleged father, considered as a random case coming to this particular paternity testing laboratory, originally had a 65% probability of being the true father, now, when we take the result of this blood typing into consideration as well, he has a 94% probability of being the true father. In practice, paternity testing is done with a special panel of genetic markers that usually results in either one or more exclusions or a very high probability of paternity. We can never prove that the alleged father is the true father, because a random man could also have the same blood type as the true father. But if enough genetic systems are typed, either an exclusion will be found or the final probability will be very close to unity. In fact it is possible, using the many new genetic systems that have been discovered in DNA (obtainable from the white cells of a blood sample), to exclude virtually everyone except the monozygotic twin of the true father.

Several points should be noted about the use of Bayes’ theorem for calculating the ‘probability of paternity’. First, when the method was initially proposed, it was assumed that \( P(D_1) = P(D_2) = 0.5 \) (and in fact, there may be many who still misguidedly make this assumption). The argument given was that the alleged father was either the true father or was not, and in the absence of any knowledge about which was the case, it would seem reasonable to give these two possibilities equal prior probabilities. To see how unreasonable such an assumption is, consider the following. If I roll a die, it will come up either a ‘six’ or ‘not a six’. Since I do not know which will happen, should I assume equal probabilities for each of these two possibilities? In fact, of course, previous experience and our understanding of physical laws suggest that the probability that a fair die will come up a ‘six’ is \( \frac{1}{6} \), whereas the probability it will come up ‘not a six’ is \( \frac{5}{6} \). Similarly, we should use previous experience and our knowledge of genetic theory to come up with a reasonable prior probability that the alleged father is the true father.

Second, although the probability of paternity obtained in this way may be perfectly valid, it may not be relevant for the particular man in question. A blood-typing laboratory may validly calculate a 99% probability of paternity for 100 different men, and exactly one of these men may not be the true father of the 100 children concerned. But if it were known, for that one man, that he could not have had access to the woman in question, the relevant prior probability for him would be 0. It was a step forward when blood-typing evidence became admissible in courts of law, but it would be a step backward if, on account of this, other evidence were ignored. The so-called probability of paternity summarizes the evidence from the blood-typing only.

Last, always remember that probabilities depend on certain assumptions. At the most basic level, a probability depends on the population of reference. Thus, when we assumed \( P(S|D_2) = 0.06 \), we were implicitly assuming a population of men in which the frequency of the B allele is 0.06 – which is appropriate for Caucasians but not necessarily for other racial groups. Had we not wished to make
this assumption, but rather that the father could have been from one of two different racial groups, it would have been necessary to assume the specific probabilities that he came from each of those racial groups. A posterior probability obtained using Bayes’ theorem also assumes we know all the possible ways that the event $B$ can occur and can specify an appropriate prior probability for each. In our example, we assumed that the true father was either the alleged father – the accused man – or a random Caucasian man. But could the true father have been a relative of the woman? Or a relative of the accused man – perhaps his brother? We can allow for these possibilities when we use Bayes’ theorem, because in the general theorem we are not limited to $k = 2$, but then we must have an appropriate prior probability for each possibility. In practice, it may be difficult to know what the appropriate prior probabilities are, even if we are sure that no relatives are involved.

The examples given above have been kept simple for instructive purposes. Nevertheless, you should begin to have an idea of the powerful tool provided by Bayes’ theorem. It allows us to synthesize our knowledge about an event to update its probability as new knowledge becomes available. With the speed of modern computers it is practical to perform the otherwise tedious calculations even in a small office setting.

**LIKELIHOOD RATIO**

If, in paternity testing, we do not know or are unwilling to assume particular prior probabilities, it is impossible to derive a posterior probability. But we could measure the strength of the evidence that the alleged father is the true father by the ratio $P(S|D_1)/P(S|D_2) = 0.5/0.06 = 8.3$, which in this particular situation is called the ‘paternity index’. It is simply the probability of what we observed (the child receiving allele $B$ from the true father) if the alleged father is in fact the father, relative to the probability of what we observed if a random Caucasian man is the father. This is an example of what is known as a likelihood ratio. If we have any two hypotheses $D_1$ and $D_2$ that could explain a particular event $S$, then the likelihood ratio of $D_1$ relative to $D_2$ is defined as the ratio $P(S|D_1)/P(S|D_2)$. The conditional probability $P(S|D_1)$ – the probability of observing $S$ if the hypothesis $D_1$ (‘the alleged father is the true father’) is true – is also called the ‘likelihood’ of the hypothesis $D_1$ on the basis of $S$ having occurred. Similarly, $P(S|D_2)$ would be called the likelihood of $D_2$. The likelihood ratio is a ratio of two conditional probabilities and is used to assess the relative merits of two ‘conditions’ ($D_1$ versus $D_2$) or hypotheses. This likelihood ratio is also a special case of what is known as a Bayes factor, which is also used to assess the relative merits of two hypotheses. These concepts have many applications in statistics. We shall discuss likelihood ratios and Bayes factors further in Chapter 8.
SUMMARY

1. The probability of event $A$ is classically defined as the number of outcomes that are favorable to $A$ divided by the total number of possible outcomes. This definition has the disadvantage that it requires one to assume that all possible outcomes are equiprobable. The frequency definition assumes one can perform a trial many times and defines the probability of $A$ as the limiting proportion, as the number of trials tends to infinity, of the trials in which $A$ occurs. The axiomatic definition of probability is simply a set of positive numbers that sum to unity.

2. A valid probability need not be the clinically relevant probability; the patient at hand may belong to a special subset of the total population to which the probability refers.

3. The probability of $A$ or $B$ occurring is given by

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B).$$

If $A$ and $B$ are mutually exclusive events, then $P(A \text{ and } B) = 0$.

4. The joint probability of $A$ and $B$ occurring is given by

$$P(A \text{ and } B) = P(A)P(B|A) = P(B)P(A|B).$$

If $A$ and $B$ are independent, $P(A \text{ and } B) = P(A)P(B)$, and conversely, if $P(A \text{ and } B) = P(A)P(B)$, then $A$ and $B$ are independent. Many different biological mechanisms can be the cause of dependence. Mutually exclusive events are never independent.

5. The conditional probability of $A$ given $B$ is given by

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}.$$

6. Bayes’ theorem states that the posterior probability that a particular $D_j$ occurred, after it is known that the event $S$ has occurred, is equal to the joint probability of $D_j$ and $S$ divided by the sum of the joint probabilities of each possible $D$ and $S$:

$$P(D_j|S) = \frac{P(D_j \text{ and } S)}{P(D_1 \text{ and } S) + P(D_2 \text{ and } S) + \cdots + P(D_k \text{ and } S)}$$

$$= \frac{P(D_j)P(S|D_j)}{P(D_1)P(S|D_1) + P(D_2)P(S|D_2) + \cdots + P(D_k)P(S|D_k)}.$$
It is assumed that we can specify a complete set of mutually exclusive ways in which \( S \) can occur, together with the prior probability of each. It does not assume that these prior probabilities are all equal.

7. When the prior probabilities of \( D_1 \) and \( D_2 \) are unknown, we can consider \( P(S|D_1)/P(S|D_2) \), the likelihood ratio of \( D_1 \) versus \( D_2 \), as a summary of the evidence provided by the event \( S \) relative to \( D_1 \) and \( D_2 \).

**FURTHER READING**


**PROBLEMS**

Problems 1–4 are based on the following: For a particular population, the lifetime probability of contracting glaucoma is approximately 0.007 and the lifetime probability of contracting diabetes is approximately 0.020. A researcher finds (for the same population) that the probability of contracting both of these diseases in a lifetime is 0.0008.

1. What is the lifetime probability of contracting either glaucoma or diabetes?
2. What is the lifetime probability of contracting glaucoma for a person who has, or will have, diabetes?
3. What is the lifetime probability of contracting diabetes for a person who has, or will have, glaucoma?
   Possible answers for Problems 1–3 are
   
   A. 0.0400
   B. 0.0278
   C. 0.0296
   D. 0.0262
   E. 0.1143
4. On the basis of the information given, which of the following conclusions is most appropriate for the two events, contracting glaucoma and contracting diabetes? They

A. are independent
B. are not independent
C. have additive probabilities
D. have genetic linkage
E. have biological variability

5. A certain operation has a fatality rate of 30%. If this operation is performed independently on three different patients, what is the probability that all three operations will be fatal?

A. 0.09
B. 0.90
C. 0.009
D. 0.027
E. 0.27

6. The probability that a certain event A occurs in a given run of an experiment is 0.3. The outcome of each run of this experiment is independent of the outcomes of other runs. If the experiment is run repeatedly until A occurs, what is the probability exactly four runs will be required?

A. 0.0531
B. 0.1029
C. 0.2174
D. 0.4326
E. 0.8793

7. A small clinic has three physicians on duty during a standard work week. The probabilities that they are absent from the clinic at any time during a regular work day because of a hospital call are 0.2, 0.1 and 0.3, respectively. If their absences are independent events, what is the probability that at least one physician will be in the clinic at all times during a regular work day? (Disregard other types of absences.)

A. 0.006
B. 0.251
C. 0.496
D. 0.813
E. 0.994
8. If two thirds of patients survive their first myocardial infarction and one third of these survivors is still alive 10 years after the first attack, then among all patients who have a myocardial infarction, what proportion will die within 10 years of the first attack? (*Hint:* Draw a tree diagram.)

A. 1/9  
B. 2/9  
C. 1/3  
D. 2/3  
E. 7/9

9. If 30% of all patients who have a certain disease die during the first year and 20% of the first-year survivors die before the fifth year, what is the probability an affected person survives past 5 years? (*Hint:* Draw a tree diagram.)

A. 0.50  
B. 0.10  
C. 0.56  
D. 0.06  
E. 0.14

10. Suppose that 5 men out of 100 and 25 women out of 10,000 are colorblind. A colorblind person is chosen at random. What is the probability the randomly chosen person is male? (Assume males and females to be in equal numbers.)

A. 0.05  
B. 0.25  
C. 0.75  
D. 0.95  
E. 0.99

11. A mother with blood type B has a child with blood type O. She alleges that a man whose blood type is O is the father of the child. What is this likelihood that the man is the true father, based on this information alone, relative to a man chosen at random from a population in which the frequency of the O allele is 0.67?

A. 0.33  
B. 1.49  
C. 2.00  
D. 0.50  
E. 0.67
12. There is a 5% chance that the mother of a child with Down syndrome has a particular chromosomal translocation, and if she has that translocation, there is a 16% chance that a subsequent child will have Down syndrome; otherwise the chance of a subsequent child having Down syndrome is only 1%. Given these facts, what is the probability, for a woman with a Down syndrome child, that her next child has Down syndrome?

A. 0.21  
B. 0.16  
C. 0.05  
D. 0.02  
E. 0.01

13. A person randomly selected from a population of interest has a probability of 0.01 of having a certain disease which we shall denote $D$. The probability of a symptom $S$, which may require a diagnostic test to evaluate its presence or absence, is 0.70 in a person known to have the disease. The probability of $S$ in a person known not to have the disease is 0.02. A patient from this population is found to have the symptom. What is the probability this patient has the disease?

A. 0.01  
B. 0.02  
C. 0.26  
D. 0.53  
E. 0.95

14. When the prior probabilities of $D_1$ and $D_2$ are unknown, the quantity $P(S|D_1)/P(S|D_2)$ is called

A. the risk of $S$ given $D_2$  
B. the correlation ratio attributable to $D_1$ and $D_2$  
C. Bayes’ theorem  
D. the joint probability of $D_1$ and $D_2$ occurring in the presence of $S$  
E. the likelihood ratio of the evidence provided by $S$ for $D_1$ relative to $D_2$

15. Let $E$ be the event ‘exposed to a particular carcinogen’, $N$ the event ‘not exposed to the carcinogen’, and $D$ the event ‘disease present’. If the likelihood ratio $P(D|E)/P(D|N)$ is 151.6, this can be considered to be a summary of the evidence that

A. disease is more likely to occur in the exposed  
B. exposure is more probable in the diseased
C. the conditional probability of disease is less than the unconditional probability
D. the conditional probability of disease is equal to the unconditional probability
E. the events $D$ and $E$ are mutually exclusive