Chapter 2
A LITTLE BIT OF PROBABILITY

The theory of probability is at bottom nothing but common sense reduced to calculus.

Pierre Simon De Le Place
Theori Analytique des Probabilites (1812–1820)

2.1 What Is Probability?

The probability of the occurrence of an event is indicated by a number ranging from 0 to 1. An event whose probability of occurrence is 0 is certain not to occur, whereas an event whose probability is 1 is certain to occur.

The classical definition of probability is as follows: if an event can occur in N mutually exclusive, equally likely ways and if n_A of these outcomes have attribute A, then the probability of A, written as $P(A)$, equals n_A/N. This is an a priori definition of probability, that is, one determines the probability of an event before it has happened. Assume one were to toss a die and wanted to know the probability of obtaining a number divisible by three on the toss of a die. There are six possible ways that the die can land. Of these, there are two ways in which the number on the face of the die is divisible by three, a 3 and a 6. Thus, the probability of obtaining a number divisible by three on the toss of a die is 2/6 or 1/3.

In many cases, however, we are not able to enumerate all the possible ways in which an event can occur, and, therefore, we use the *relative frequency definition of probability.* This is defined as the number of times that the event of interest has occurred divided by the total number of trials (or opportunities for the event to occur). Since it is based on previous data, it is called the *a posteriori definition of probability.*

For instance, if you select at random a white American female, the probability of her dying of heart disease is .00287. This is based on the finding that per 100,000 white American females, 287 died of coronary heart disease (estimates are for 2001, National Center for Health Statistics[7]). When you consider the probability of a white American female who is between ages 45 and 64, the figure drops to .00088 (or 88 women in that age group out of 100,000), and when you consider women 65 years and older, the figure rises to .01672 (or 1672 per 100,000). For white men 65 or older it is .0919 (or 9190 per 100,000). The two important points are (1) to determine a probability, *you must specify the population to which you refer*, for example, all white females, white males between 65 and 74, nonwhite females between 65 and 74, and so on; and (2) the *probability figures are constantly revised* as new data become available.

This brings us to the notion of *expected frequency*. If the probability of an event is P and there are N trials (or opportunities for the event to occur), then we can expect that the event *will* occur $N \times P$ times. It is necessary to remember that probability "works" for large numbers. When in tossing a coin we say the probability of it landing on heads is .50, we mean that in many tosses half the time the coin will land heads. If we toss the coin ten times, we may get three heads (30%) or six heads (60%), which are a considerable departure from the 50% we expect. But if we toss the coin 200,000 times, we are very likely to be close to getting exactly 100,000 heads or 50%.

Expected frequency is really the way in which probability "works." It is difficult to conceptualize applying probability to an individual. For example, when TV announcers proclaim there will be say, 400 fatal accidents in State X on the Fourth of July, it is impossible to say whether any individual person will in fact have such an accident, but we can be pretty certain that the number of such accidents will be very close to the predicted 400 (based on probabilities derived from previous Fourth of July statistics).

2.2 Combining Probabilities

There are two laws for combining probabilities that are important. First, if there are *mutually exclusive events* (i.e., if one occurs, the other cannot), the probability of either one or the other occurring is the *sum* of their individual probabilities. Symbolically,

$$P(A\ or\ B) = P(A) + P(B)$$

An example of this is as follows: the probability of getting either a 3 or a 4 on the toss of a die is 1/6 + 1/6 = 2/6.

A useful thing to know is that the sum of the individual probabilities of all possible mutually exclusive events must equal 1. For example, if A is the event of winning a lottery, and not A (written as $\overline{A}$), is the event of not winning the lottery, then $P(A) + P(\overline{A}) = 1.0$ and $P(\overline{A}) = 1 - P(A)$.

Second, if there are two independent events (i.e., the occurrence of one is not related to the occurrence of the other), the joint probability of their occurring together (jointly) is the *product* of the individual probabilities. Symbolically,

$$P(A\ and\ B) = P(A) \times P(B)$$

An example of this is the probability that on the toss of a die you will get a number that is both even and divisible by 3. This probability is equal to $1/2 \times 1/3 = 1/6$. (The only number both even and divisible by 3 is the number 6.)

The joint probability law is used to test whether events are independent. If they are independent, the product of their individual probabilities should equal the joint probability. If it does not, they are not independent. It is the basis of the chi-square test of significance, which we will consider in the next section.

Let us apply these concepts to a medical example. The mortality rate for those with a heart attack in a special coronary care unit in a certain hospital is 15%. Thus, the probability that a patient with a heart attack admitted to this coronary care unit will die is .15 and that he will survive is .85. If two men are admitted to the coronary care unit on a particular day, let A be the event that the first man dies and let B be the event that the second man dies.

The probability that both will die is

$$P(A\ and\ B) = P(A) \times P(B) = .15 \times .15 = .0225$$

We assume these events are independent of each other so we can multiply their probabilities. Note, however, that the probability that either one *or* the other will die from the heart attack is *not* the sum of their probabilities because these two events are not mutually exclusive. It is possible that both will die (i.e., both *A* and *B* can occur).

To make this clearer, a good way to approach probability is through the use of Venn diagrams, as shown in Figure 2.1. Venn diagrams consist of squares that represent the universe of possibilities and circles that define the events of interest.

In diagrams 1, 2, and 3, the space inside the square represents all *N* possible outcomes. The circle marked A represents all the outcomes that constitute event *A*; the circle marked B represents all the outcomes that constitute event *B*. Diagram 1 illustrates two mutually exclusive events; an outcome in circle A cannot also be in circle B. Diagram 2 illustrates two events that can occur jointly: an outcome in circle A can also be an outcome belonging to circle B. The shaded area marked AB represents outcomes that are the occurrence of both *A and B*. The diagram 3 represents two events where one (*B*) is a subset of the other (*A*); an outcome in circle B must also be an outcome constituting event *A*, but the reverse is not necessarily true.

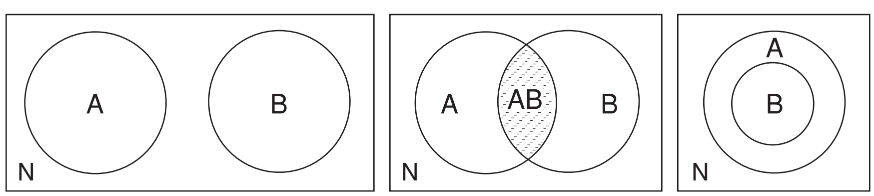

Figure 2.1

It can be seen from diagram 2 that if we want the probability of an outcome being either *A or B* and if we add the outcomes in circle A to the outcomes in circle B, we have added in the outcomes in the shaded area twice. Therefore, we must subtract the outcomes in the shaded area (A *and* B) also written as (AB) once to arrive at the correct answer. Thus, we get the result

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

2.3 Conditional Probability

Now let us consider the case where the chance that a particular event happens is dependent on the outcome of another event. The probability of A, given that B has occurred, is called the conditional probability of A given B, and is written symbolically as $P(A|B)$. An illustration of this is provided by Venn diagram 2. When we speak of conditional probability, the denominator becomes all the outcomes in circle B (instead of all N possible outcomes) and the numerator consists of those outcomes that are in that part of A which also contains outcomes belonging to B. This is the shaded area in the diagram labeled AB. If we return to our original definition of probability, we see that

$$P(A \mid B) = \frac{n_{AB}}{n_B}$$

(the number of outcomes in both A *and* B, divided by the total number of outcomes in B).

If we divide both numerator and denominator by N, the total number of *all* possible outcomes, we obtain

$$P(A \mid B) = \frac{n_{AB}/N}{n_B/N} = \frac{P(A \text{ and } B)}{P(B)}$$

Multiplying both sides by $P(B)$ gives the *complete* multiplicative law:

$$P(A \text{ and } B) = P(A \mid B) \times P(B)$$

Of course, if A and B are independent, then the probability of A given B is just equal to the probability of A (since the occurrence of B does not influence the occurrence of A) and we then see that

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

2.4 Bayesian Probability

Imagine that M is the event "loss of memory," and B is the event "brain tumor." We can establish from research on brain tumor patients the probability of *memory loss given a brain tumor,* $P(M|B)$. A clinician, however, is more interested in the probability of *a brain tumor, given that a patient has memory loss,* $P(B \mid M)$.

It is difficult to obtain that probability directly because one would have to study the vast number of persons with memory loss (which in most cases comes from other causes) and determine what proportion of them have brain tumors.

Bayes' equation (or Bayes' theorem) estimates $P(B \mid M)$ as follows:

$$P(\text{brain tumor, given memory loss}) = \frac{P(\text{memory loss, given brain tumor}) \times P(\text{brain tumor})}{P(\text{memory loss})}$$

In the denominator, the event of "memory loss" can occur either among people with brain tumor, with probability = $P(M \mid B)\ P(B)$, or among people with no brain tumor, with probability = $P(M \mid \bar{B})P(\bar{B})$. Thus,

$$P(B \mid M) = \frac{P(M \mid B)P(B)}{P(M \mid B)P(B) + P(M \mid \bar{B})P(\bar{B})}$$

The overall probability of a brain tumor, $P(B)$ is the "a priori probability," which is a sort of "best guess" of the prevalence of brain tumors.

2.5 Odds and Probability

When the odds of a particular horse *losing* a race are said to be 4 to 1, he has a 4/5 = .80 probability of losing. To convert an odds statement to probability, we add 4 + 1 to get our denominator of 5. The odds of the horse *winning* are 1 to 4, which means he has a probability of winning of 1/5 = .20.

$$\textit{The odds in favor of } A = \frac{P(A)}{P(not\ A)} = \frac{P(A)}{1 - P(A)}$$

$$P(A) = \frac{odds}{1 + odds}$$

The odds of drawing an ace = 4 (aces in a deck) to 48 (cards that are not aces) = 1 to 12; therefore, P(ace) = 1/13. The odds *against* drawing an ace = 12 to 1; P(Not Ace) = 12/13.

In medicine, odds are often used to calculate an *odds ratio.* An odds ratio is simply the ratio of two odds. For example, say that in a particular study comparing lung cancer patients with controls, it was found that the odds of being a lung cancer case for people who smoke were 5 to 4 (5/4) and the odds of having lung cancer for nonsmokers was 1 to 8 (1/8), then the odds ratio would be

$$\frac{5/4}{1/8} = \frac{5 \times 8}{4 \times 1} = \frac{40}{4} = 10$$

An odds ratio of 10 means that the odds of being a lung cancer case is 10 times greater for smokers than for nonsmokers.

Note, however, that we cannot determine from such an analysis what the probability of getting lung cancer is for smokers, because in order to do that we would have to know how many people out of all smokers developed lung cancer, and we haven't studied all smokers; all we do know is how many out of all our lung cancer cases were smokers. Nor can we get the probability of lung cancer among nonsmokers, because we would have to a look at a population of nonsmokers and see how many of them developed lung cancer. All we do know is that smokers have 10-fold greater odds of having lung cancer than nonsmokers.

More on this topic is presented in Section 4.12.

2.6 Likelihood Ratio

A related concept is the likelihood ratio (LR), which tells us how likely it is that a certain result would arise from one set of circumstances in relation to how likely the result would arise from an opposite set of circumstances.

For example, if a patient has a sudden loss of memory, we might want to know the likelihood ratio of that symptom for a brain tumor, say. What we want to know is the likelihood that the memory loss arose out of the brain tumor *in relation to* the likelihood that it arose from some other condition. The likelihood ratio is a ratio of conditional probabilities.

$$
\begin{aligned}
LR &= \frac{P(\textit{memory loss, given brain tumor})}{P(\textit{memory loss, given no brain tumor})} \\
&= \frac{P(M \textit{ given } B)}{P(M \textit{ given not } B)}
\end{aligned}
$$

Of course to calculate this LR we would need to have estimates of the probabilities involved in the equation, that is, we would need to know the following: among persons who have brain tumors, what proportion have memory loss, and among persons who don't have brain tumors, what proportion have memory loss. It may sometimes be quite difficult to establish the denominator of the likelihood ratio because we would need to know the prevalence of memory loss in the general population.

The LR is perhaps more practical to use than the Bayes' theorem, which gives the probability of a particular disease given a particular symptom. In any case, it is widely used in variety of situations because it addresses this important question: If a patient presents with a symptom, what is the likelihood that the symptom is due to a particular disease *rather than* to some other reason than this disease?

2.7 Summary of Probability

Additive Law:

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

If events are mutually exclusive: $P(A \text{ or } B) = P(A) + P(B)$.

Multiplicative Law:

$$P(A \text{ and } B) = P(A \mid B) \times P(B)$$

If events are independent: $P(A \text{ and } B) = P(A) \times P(B)$.

Conditional Probability:

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

Bayes' Theorem:

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

Odds of A:

$$\frac{P(A)}{1 - P(A)}$$

Likelihood Ratio:

$$\frac{P(A \mid B)}{P(A \mid \bar{B})}$$