

relation of each with IQ. For persons of the same IQ, there is little correlation between speed and comprehension. Hence, since an increase in reading speed will not increase IQ, it would be expected to have little effect on comprehension.

6. (a) $\bar{Y}_i = .00186V_i + 2.51$
 (b) $r_{YQ,V} = -.013$, $r_{YG,V} = -.122$; G will be selected since its absolute value is greatest

(c) $\hat{z}_{Y_i} = .379z_{V_i} - .121z_{G_i}$

(d) .258

(e) $\bar{Y}_i = .00207V_i - .165G_i + 3.133$

(f) 4.06

(g) .34

(h) .12 or 12%

(i) yes, .35

7. $\hat{r}^2 = 1 - 701.4/1,041.6 = .3266$; hence,
 $\hat{r} = \sqrt{.3266} = .57 = r$

9

PROBABILITY

9.1 INTRODUCTION

Researchers often attempt to generalize from their observations. They make a tacit assumption that the set of data has some generalizability—if they gathered more data tomorrow, it would reflect the same general trend. Inferences differ in their likelihood of being correct all the way from “extremely unlikely” to “almost certain.” From the standpoint of logic, all inferences contain uncertainty. Statisticians have developed methods that assign probabilities to inferences. Inferential reasoning is a principal method of science. The language of everyday life—“extremely unlikely” or “almost certain”—lacks precision.

Statisticians do not completely agree as to how to assign probabilities to statements or how to choose to which statements to assign probabilities. Nonetheless, their preference for objectivity and quantification springs from values regarding the nature and methods of science.

The remaining chapters deal with assigning a probability value to an inference. The methods that statisticians have developed allow one to state, for example, “There is a positive relationship between IQ and grade-point average, and the probability of obtaining such a large r by chance is very small—only .01 if there really is no correlation in the population.” Only certain basics of probability theory can be addressed in a single chapter; probability theory is a large and complex body of knowledge. The finer points in probability theory are not needed for using and interpreting the probability statements used in statistical inference. An intuitive understanding of probability is, however, necessary to interpret the statistics of hypothesis testing and interval estimation.

9.2 PROBABILITY AS A MATHEMATICAL SYSTEM

Probability can be viewed as a system of definitions and operations pertaining to a sample space. The idea of a *sample space* is basic. Every probability statement is related to a

sample space of some sort; indeed, statements of probability are statements about sample spaces.¹

A sample space can be defined as a set of points. These points can represent persons, businesses, cities, schools, et cetera. An *event* is an observable happening like the appearance of heads when a coin is flipped, or that a person is watching television. There are usually many points in the sample space, each of which is an example of an event. For instance, the sample space may be a set of six white and three black balls in an urn. This sample space has nine points. An event might be "A ball is white." This event has six sample-space points. How many points in the sample space does the event "A ball is black" have? The event "A ball in this urn is red" has no sample points. "A ball in this urn is either white or black" is also an example of an event. Notice that many different events can be defined on the same sample space.

A statement of probability is made about the relative frequency of an event that is associated with a sample space. A capital letter, A, B, C, \dots , will stand for an event; the "probability of the event A " will be denoted by $P(A)$.

Definition: The probability of the event A , $P(A)$, is the ratio of the number of sample points that are examples of A to the total finite number of sample points in the sample space, assuming all sample points are equally likely.

Let A be the event "a 3 face of a die," where the sample space is the set of the six faces of a die. How many sample points are examples of the event A ? Obviously, only one. The total number of sample points is six. Hence, the probability of the event A (3) is:

$$P(A) = \frac{\text{Number of Examples of } A}{\text{Total Number of Sample Points}} \quad (9.1)$$

$$P(A) = \frac{1}{6}$$

If B is the event "an even numeral" in this die-tossing example, find $P(B)$. B can be restated as "a 2 or 4 or 6." Since B consists of three points in the sample space of six points, $P(B) = 3/6 = 1/2$. What is $P(C)$ if C is the event 7? $P(C)$ is $0/7 = 0$, because 7 is not in the sample space of this problem. If D is the event "an even or odd numeral," what is $P(D)$? The answer is $6/6 = 1$.

Suppose there is an urn that has four white balls in it and a finite, but unspecified, number of black balls. The probability of an event cannot be determined. A probability statement can be made only when the sample space is defined completely.

The definition of the probability of an event can be expressed using an alternative approach. Consider a sample space composed of a specified number of sample points. Denote each of the sample points by " a_i ": a_1, a_2, \dots, a_n . Every event that is defined within the sample space is composed of a related set of sample points.

¹The notion of a sample space is actually a relatively recent development in probability theory, dating back only to the 1920s.

Definition: A probability function is a rule of correspondence that associates with each event A in the sample space, a number $P(A)$ such that:

1. For any event A , $1 \geq P(A) \geq 0$.
2. The sum of the probabilities for all distinct events is 1.
3. If A and B are mutually exclusive events, that is, have no sample points in common, the $P(A \text{ or } B) = P(A) + P(B)$.

If it is assumed that the probability of every elementary event a_i is $1/N$, where N is the total number of sample points, then the probability of the event A that is composed of r sample points² is:

$$P(A) = \frac{1}{N} + \frac{1}{N} + \dots + \frac{1}{N} = \frac{r}{N} \quad (9.2)$$

The probability of event A , $P(A)$, is the ratio of the number, r , of sample points that are examples of A to the total number of sample points, N , that is, r/N .

Both routes bring us to the same definition for $P(A)$. While the second definition might be the preference of the mathematician, the first definition of $P(A)$ is more intuitive.

Combining Probabilities

Suppose an urn contains five red, three white, and two black balls. Three events might be of interest: (1) A , a ball is red, (2) B , a ball is white, or (3) C , a ball is black. These three events are mutually exclusive; each sample point is an example of one and only one event.

The question arises, "What is the probability that a ball is red or white?" This event, the *union* of A and B , shall be denoted by the symbol $A \cup B$ and its probability by $P(A \cup B)$.

9.3 FIRST ADDITION RULE OF PROBABILITIES

When the events A and B are mutually exclusive, the probability of either A or B , $P(A \cup B)$, is:

$$P(A \cup B) = P(A) + P(B) \quad (9.3)$$

From the example above, we have:

$$P(A \cup B) = P(A) + P(B) = \frac{5}{10} + \frac{3}{10} = \frac{8}{10} \text{ or } .8$$

²The symbol conventionally used in probability theory for this purpose is r , not to be confused with the correlation coefficient.

Find $P(A \cup C)$. Since events A and C are mutually exclusive:

$$P(A \cup C) = P(A) + P(C) = .5 + .2 = .7.$$

Similarly, the value of $P(B \cup C) = P(B) + P(C) = .3 + .2 = .5$.

Non-mutually Exclusive Events

In some sample spaces, two events are not mutually exclusive: a single sample point may be an example of both events A and B . A playing card can be both an ace and a diamond. Consider the possible outcomes (heads or tails) of flipping a fair coin three times in a row, or three fair coins once. The eight possible outcomes make up the sample space:

1. HHH
2. HHT
3. HTH
4. HTT
5. THH
6. THT
7. TTH
8. TTT

Each of the eight outcomes is equally likely, that is, each has probability $1/8$. What is the probability of heads on the first flip? The answer is $4/8$ or $1/2$. What is the probability of heads on flips 1 and 2? The answer is $2/8$ or $1/4$.

Now define two events, A and B , using the sample space just defined:

A : Heads on flips 1 and 2

B : Heads on flips 2 and 3

The sample points that are examples of event A are the first two events (HHH and HHT) in the sample space. The first and fifth outcomes (HHH and THH) are the sample points corresponding to event B . The symbol $A \cap B$ shall denote the new event, the *intersection* of A and B . (Note the symbols \cup and \cap are analogous to the words *or* and *and*.) In the example, $A \cap B$ is the event "heads on flips 1 and 2 and heads on flips 2 and 3". Since all the sample points are equally likely, the probability of the event $A \cap B$ is:

$$P(A \cap B) = \frac{\text{Number of Sample Points that Are Examples of } A \cap B}{\text{Total Number of Sample Points}} \quad (9.4)$$

The total number of sample points is 8. Only one sample point, HHH, is an example of the event A "heads on flips 1 and 2"; and also an example of event B "heads on flips 2 and 3". So the probability of the event $A \cap B$ is $1/8$.

Look back at the first addition rule of probabilities and notice the condition that the two events A and B are mutually exclusive. In the example just discussed, A and B were *not* mutually exclusive. The outcome HHH was an example of both events A and B . But what is the probability of A or B , $P(A \cup B)$, when A and B are not mutually exclusive? The "Second Addition Rule of Probabilities" is needed to answer this question.

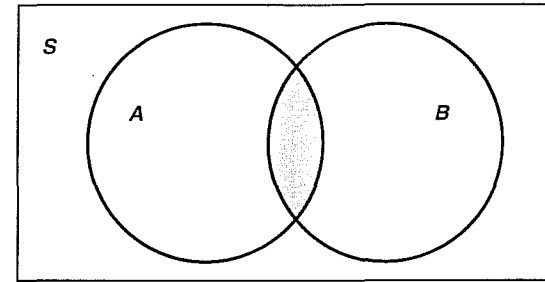


FIGURE 9.1

Venn diagram of the intersecting events A and B in the sample space S .

9.4 SECOND ADDITION RULE OF PROBABILITIES

The probability of either event A , or event B , or both is expressed as:

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

The Venn diagram in Figure 9.1 is a graphic portrayal of this situation, and should help clarify the meaning of the term $P(A \cap B)$. The events A and B are not mutually exclusive, that is, they have some sample points in common in the sample space S . The probability of event A is represented by the area of circle A ; the probability of event B is represented by the area of circle B . The probability of A or B , or *both*, is that area of S that is inside the boundaries of both A and B . The shaded portion in Figure 9.1 is that set of sample points in both events A and B , that is, those points in the *intersection* $A \cap B$.

How does one find the entire area covered by A and B ? First, find the area of A that is not shared by B . Add to it the area of B not shared by A , and then add the area common to both—the intersection of A and B :

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

The first two terms following the equal sign, $P(A) - P(A \cap B)$, give the area of A minus the area in common with B . $P(B) - P(A \cap B)$ gives the area of B minus its area in common with A . The desired area is found by adding in the area common to A and B , $P(A \cap B)$. The previous equation simplifies to

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad (9.5)$$

Equation 9.5 establishes the *second addition rule of probabilities*. Notice that if one had simply added $P(A)$ and $P(B)$ to find $P(A \cup B)$, the portion in common to A and B , $P(A \cap B)$, would contribute twice to the sum, because A and B are not mutually exclusive areas. The intersection must be included only once; consequently it must be subtracted once as shown in Equation 9.5.

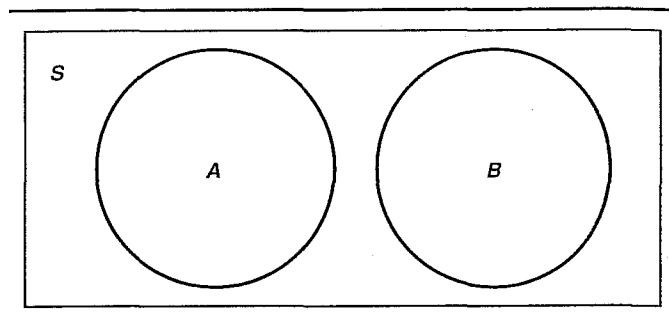


FIGURE 9.2

Venn diagram of the mutually exclusive events A and B in the sample space S .

Notice the first addition rule of probabilities (Equation 9.3) is just a special case of the second rule (Equation 9.5), that is, in the case when $P(A \cap B) = 0$. If A and B are mutually exclusive events in S , then they do not overlap. See Figure 9.2, where there is no area in common to A and B ; hence $P(A \cap B) = 0$. More generally, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. If A and B are mutually exclusive, then $P(A \cap B) = 0$. Therefore, if A and B are mutually exclusive:

$$P(A \cup B) = P(A) + P(B) - 0 = P(A) + P(B)$$

In much the same manner as the mathematical systems of geometry and algebra, probability theory can be developed from a small set of axioms and definitions. Also in the same manner, probability theory can serve as a model for what is going on in a certain class of events in the world around us. James Bernoulli (1654–1705) was the first to relate probability statements to physical events.³ An example of the application of a formal probability statement to an actual set of actions will illustrate the relationship between theory and application.

Suppose an urn contains four white and six black balls. The balls are identical in size, shape, and weight and thoroughly mixed so that if one were to reach in and pull one out, it is equally likely that any one of the ten balls would be selected; each ball has one chance in ten of being chosen. A ball is taken out and its color is recorded. The ball is returned to the urn, the balls in the urn are stirred thoroughly, and the act is repeated under the same conditions. This act is performed many times, say, 10,000. After the 10,000th drawing of a ball, suppose a count is taken of the number of times a white ball was drawn. Intuition says that the ratio of the number of times a white ball is drawn to 10,000 will be very close to $4/10 = .4$. It is unlikely the ratio is exactly $4/10$ but it will be very close.

If the ten balls are regarded as a sample space, and if A is the event “a ball is white,” then $P(A)$ is exactly $.4$. The question arises, “Will the formal probability of an event as calculated from theory correspond closely to the relative frequency of the occurrence of the

³The “Bernoulli distribution” is another name for the binomial distribution that you may have long forgotten from high school algebra. This distribution has only two classes of independent events in the sample space, for example, heads or tails, hit or miss, event A or not-event A .

event?” The answer to this question is the key to the relationship of probability theory and its application. The answer is yes when the underlying assumptions are met. Suppose an event A either does, or does not, occur on every trial of an act. The probability that A will occur, $P(A)$, is the same for all trials of the act. For example, the act may be flipping a fair coin, A may be the event “heads,” and it is assumed that the probability of heads ($1/2$) is the same from one flip to the next. It is also assumed that every trial is independent of, in no way affected by, every other trial. Now after N trials of the act, the proportion of times A has occurred is p . It can be proved that p gets closer and closer to $P(A)$ as N becomes larger and larger. The proportion of times A occurs can be made closer and closer to $P(A)$ (the probability calculated from the sample space) by performing the act an increasing number of times. So $P(A)$ tells what will happen *in the long run* if the actions are actually performed under the conditions laid down previously.

The preceding paragraph is a statement of the *law of large numbers*. The law of large numbers is important for the application of probability, of which statistical inference is one such application. The law of large numbers is closely related to the statistical property of consistency, encountered previously in Chapter 5.

9.5 MULTIPLICATION RULE OF PROBABILITIES

There exists a multiplication rule for probabilities that is important for statistical inference. Suppose a coin is flipped five times in a row. Assume that the probability of heads is $1/2$ on each flip and that the flips are independent (the outcomes are uncorrelated and have no influence on one another). The multiplicative rule for probabilities states that the probability of getting five straight heads is $(1/2)(1/2)(1/2)(1/2)(1/2) = 1/32$. A general statement of the rule follows:

Multiplication Rule of Probabilities: The probability that A , which has probability $P(A)$, will occur r times in r independent trials is:

$$P(A) \cdot P(A) \cdot \dots \cdot P(A) = P(A)^r \quad (9.6)$$

Illustrations

Assume that the probability that your next child will be a girl is $1/2$ or $.5 = P(G)$. Since the sex of one child has no effect on the sex of any subsequent child, the probabilities are independent. Thus, the probability of five consecutive girls (or boys) from Equation 9.6 is:

$$P(G)^5 = \left(\frac{1}{2}\right)^5 = \frac{1}{32} = .03125 \approx .03$$

In other words, of every 32 families that have five children, one (or about three of 100) would be expected to have all girls. Suppose a sixth child is now expected, what are the odds that it is girl? The odds are 1 in 2; past events that are independent have no influence on future events. To believe otherwise is to believe in the “Gambler’s Fallacy” (Section 9.12), and be in jeopardy of wasting a great deal of your money.

⁴Actually the probability is slightly less than $.49$, but we will keep the numbers simple and use $.5$.

TABLE 9.1 Sample Space of Outcomes of Tossing a Pair of Dice (One Red, One Green)

R,G	R,G	R,G	R,G	R,G	R,G
1,1	2,1	3,1	4,1	5,1	6,1
1,2	2,2	3,2	4,2	5,2	6,2
1,3	2,3	3,3	4,3	5,3	6,3
1,4	2,4	3,4	4,4	5,4	6,4
1,5	2,5	3,5	4,5	5,5	6,5
1,6	2,6	3,6	4,6	5,6	6,6

Suppose a toddler is attracted to your computer and pecks keys randomly on the keyboard. What is the probability that the result of the first six characters are M-O-T-H-E-R? For simplicity, assume that there are 100 keys on your keyboard; the probability of a correct peck is, therefore, .01. Since there are six independent events, the probability that the first six random pecks will result in "mother" is $(.01)^6 = .000000000001$ or only one chance in a trillion⁵ (1/1,000,000,000,000)!

The following examples illustrate the probability rules developed so far. Consider a roll of a pair of dice; one die is red and the other is green. The sample space of possible outcomes has 36 points, as shown in Table 9.1. Suppose event A is "a 1 on the red die" and event B is "a 1 on the green die." $P(A \cap B)$ is found by dividing the number of sample points that are examples of $A \cap B$ (both A and B) by the total number of sample points (36). Verify that the probability that both A and B occur is equal to $P(A) \cdot P(B)$, $(1/6)(1/6) = 1/36$.

Find $P(A \cup B)$, the probability of event A or event B , remembering that $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. $P(A \cap B)$ is $1/36$ because only the point "1,1" is common to A and B ; thus, $P(A \cup B) = 6/36 + 6/36 - 1/36 = 11/36$.

Two events are *independent* if and only if $P(A \cap B) = P(A) \cdot P(B)$. Independence is an important concept in statistics and probability, and statistical independence is a much used concept in subsequent chapters.

9.6 CONDITIONAL PROBABILITY

If $P(A)$ and $P(B)$ are known, the *conditional* probability of B given A , denoted as $P(B|A)$, is:

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \quad (9.7)$$

⁵In scientific notation this is 1.00×10^{-12} . A 17-key sequence, such as I-L-O-V-E-S-T-A-T-I-S-T-I-C-S, has a probability of only $(.01)^{17}$ or 1 in 10^{34} ; if one entered a random 17-key sequence once every second, this particular sequence would be expected only once every 317,000,000,000,000,000,000,000 (or 3.17×10^{26}) years! Scientists have estimated the age of the earth at about 5 billion (5×10^9) years; thus, the probability that this 17-key sequence would have occurred even once since the beginning of the earth is $(5 \times 10^9)/(3.17 \times 10^{26}) = 6.34 \times 10^{-17}$ (less than 1 chance in 100,000,000,000,000,000; $p < .000000000000000001$)! If one billion individuals were each striking 17 keys randomly each second, the probability remains infinitesimal (1 in 100,000,000). Contrast the probability of the simple I-L-O-V-E-S-T-A-T-I-S-T-I-C-S to the incredibly greater complexity of a DNA molecule!

In Figure 9.1, note that the conditional probability $P(B|A)$ (read "the probability of B given A ") is represented by the ratio of the area $A \cap B$ to the area A . In the example in Section 9.5: $P(A) = 1/6$ and $P(A \cap B) = 1/36$; thus using Equation 9.7, the probability of B given A is $1/36 \div 1/6 = 1/6$. Given A (1 on the red die), the probability of B (1 on the green die) is $1/6$.

Suppose that the sample space is American adults and B is "a woman" and A is "a college graduate." Given that a college graduate is selected (A), what is the probability that it is a woman, that is, what is $P(B|A)$? Currently, the probability of selecting a college graduate is approximately $.26 = P(A)$, and the probability of selecting a woman college graduate is $.115 = P(A \cap B)$. Given A , that a college graduate is selected, the probability of B , a woman, is

$$P(B|A) = \frac{.115}{.26} = .44$$

Notice that $P(B|A)$ is not equal to $P(A|B)$. In this example, given B , a woman, $P(B) = .5$, the probability of A , a college graduate, is:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{.115}{.5} = .23$$

9.7 BAYES'S THEOREM

Equation 9.7 is the simplest version of Bayes's theorem,⁶ a theorem that describes the relationship among various conditional probabilities. Equation 9.7 may be expressed alternatively as:

$$P(A|B) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|\bar{B})P(\bar{B})} \quad (9.8)$$

where \bar{B} is read "not B ."

Let event B be "an automobile accident during the next year" and event A be "a course in driver education." What is the probability of event B (an accident in the next year), given A (driver education), that is, what is $P(B|A)$? Assume that $P(B) = .1$; thus $P(\bar{B}) = 1 - P(B) = .9$. Also, assume that $P(A|B)$, the probability of having driver education given that a person has had an accident, is $.50$; and that $P(A|\bar{B})$, the probability of having driver education given that a person has *not* had an accident, is $.7$. From Equation 9.8, the probability of an accident (B), given driver education (A) is:

$$P(B|A) = \frac{(.5)(.1)}{(.5)(.1) + (.7)(.9)} = .0735$$

Thus, the probability of an accident has been reduced by 26.5%, from $.10$ to $.0735$, given the person had driver education.

⁶The theorem is named for its originator, the English clergyman and mathematician Thomas Bayes (1702–1761), who first used probability inductively and established a mathematical basis for probabilistic inference.

TABLE 9.2 Permutations of Three Letters: A, B, and C

1st letter	2nd letter	3rd letter	Permutation
A	B	C	1. ABC
	C	B	2. ACB
B	A	C	3. BAC
	C	A	4. BCA
C	A	B	5. CAB
	B	A	6. CBA

Bayes's theorem provides exact results, providing the prior probabilities, the probabilities entered into Equation 9.8, are accurate. Obviously, the practical difficulty in the application of the theorem lies in knowing the prior probabilities. These probabilities have often been viewed as degrees of belief, or personal probabilities. The topic has occasioned much controversy among statisticians who favor a strict relative-frequency interpretation of probability and those who would allow for a more subjective interpretation. Discriminant analysis (Section 8.30) makes use of prior probabilities and Bayes's theorem in determining probabilities that a given case falls in a given category, given a prior probability of being in that category.

9.8 PERMUTATIONS

Two additional concepts that crop up repeatedly in probability illustrations are permutations and combinations (Section 9.9).

A *permutation* of a set of objects (the letters A, B, and C, for example) is an arrangement of them in which order is considered. A different ordering of the objects is a different permutation. How many different permutations or orderings are there of the letters A, B, and C? To find out, one can set about the task of writing them down and counting them, as shown in Table 9.2.

The first letter can be either A, B, or C. Suppose it is A, the top third of Table 9.2. If the first letter is A, the second letter can be either B or C. If the second letter is B, then the third letter must be C. So ABC is one possible permutation. There are three possible letters for the first position. After one letter is assigned to the first position, there are two possible letters for the second position. Hence, the number of possible permutations of the three letters A, B, and C is $(3)(2)(1) = 6$.

If there are N distinct objects, one can make $N(N-1)(N-2) \dots (2)(1)$ different permutations of them. This product can be denoted simply by $N!$ (read " N factorial"). $N!$ is the product of the numbers from 1 through N and equals the number of permutations of N distinct objects.⁷ ($0!$ is defined mathematically to equal 1.)

The value of $N!$ increases dramatically as N increases. Would you work a year for 12! pennies? ($12!$ pennies = 479,001,600¢ = \$4,790,016). To illustrate the incredible size of $12!$, imagine that you have one dozen eggs in a carton and that you want to form every

⁷Many calculators have a factorial (!) key that provides almost instant answers to $N!$ questions.

possible permutation, or arrangement, of them in the egg carton. Assume you can make a new arrangement in ten seconds, which you do continuously during your eight hour working day. If you keep at this job five days a week fifty-two weeks a year, you would require approximately 640 years to make every possible arrangement! If you and each successive generation of your descendants donate 50 years to this task, it would be finished by your great, great, great, great, great, great, great, great, great, great, great grandchild. We suggest you take our word for it.

On the Wechsler intelligence tests, an examinee must put five cartoon pictures in the correct chronological order. What is the probability that an examinee will arrive at the correct order by chance? Since $5! = 120$, the probability of a "lucky guess" is only $1/120$.

9.9 COMBINATIONS

The concept of combinations arises when one is selecting some number of objects r from a set of N objects. A *combination* of objects is a distinct set of objects in which order is not considered. When $r = N$, that is, N objects are selected from N objects, all the objects are selected and there is only one combination (although there are $N!$ permutations). If $r = 1$, one object is selected from N objects, and there are N combinations. The problem is to find a general expression for the number of combinations that exist when r things are selected from N things.

Consider four objects, A, B, C, and D. How many different combinations can be made by selecting two letters at a time from these four? The answer is six: AB, AC, AD, BC, BD, CD. Notice that for combinations, order is not considered: AB is one combination, and BA is the same combination. (See the first two columns of Table 9.2, where the six permutations form three combinations.)

Suppose r objects are being selected from N objects. How many different combinations are there? For the time being, regard order as important, and then later combine all the sets that are different only because of order. If r objects are being selected from N , then there are N choices for the first object, $(N-1)$ choices for the second, $(N-2)$ for third, $(N-3)$ for the fourth, and so on until there are $(N-r+1)$ choices for the r th object. So, the *total of different permutations* of r objects from N objects, where order is considered, is equal to:

$$N(N-1)(N-2) \dots (N-r+1) \quad (9.9)$$

How many permutations are there for $N = 4$ objects, taking $r = 3$ at a time? Using Equation 9.9, the answer is found to be $(4)(3)(2) = 24$ permutations. There are r terms in this product corresponding to the r objects selected. However, for each unique combination of r objects, there are $r!$ permutations; in the example, $r = 3$ and $r! = 3! = 6$. Hence, the number of combinations of r objects selected from N objects, ignoring the order among the r objects, is the number of permutations of the r objects selected from N objects (Equation 9.9) divided by the number of permutations within a combination.

$$\binom{N}{r} = \frac{N(N-1)(N-2) \dots (N-r+1)}{r!} \quad (9.10)$$

The expression to the left of the equal sign in Equation 9.10 is read "the number of combinations of r things taken from N things."

It can be shown⁸ that the number of permutations of r objects taken from N objects is:

$$N(N-1)(N-2)\dots(N-r+1) = \frac{N!}{(N-r)!} \quad (9.11)$$

When this substitution is made in Equation 9.10, the number of combinations of N things taken r at a time is given by:

$$\binom{N}{r} = \frac{N!}{r!(N-r)!} \quad (9.12)$$

Examples

How many different combinations are there of $r = 3$ things taken from $N = 5$ things?

$$\binom{5}{3} = \frac{5!}{3!(5-3)!} = \frac{5!}{3!2!} = \frac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{(3 \cdot 2 \cdot 1)(2 \cdot 1)} = \frac{5 \cdot 4}{2 \cdot 1} = \frac{20}{2} = 10$$

Note from Equation 9.9 that there are $(5)(4)(3)$ permutations of five things taken three at a time. Each combination has $r! = 3! = 6$ permutations; thus, the sixty permutations represent 60/6 combinations. Each of the ten combinations of five things taken three at a time has $(3)(2)(1) = 6$ permutations.

Ten persons are eligible to serve on a committee. The committee must be composed of only five persons. How many different five member ($r = 5$) committees could be formed from the ten available persons ($N = 10$)? From Equation 9.12:

$$\binom{10}{5} = \frac{10!}{5!(10-5)!} = \frac{10!}{5!5!} = \frac{10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{(5 \cdot 4 \cdot 3 \cdot 2 \cdot 1)(5 \cdot 4 \cdot 3 \cdot 2 \cdot 1)} = \frac{30,240}{120} = 252$$

9.10 BINOMIAL PROBABILITIES

When we have a Bernoulli-type (two classes of independent events) sample space and the probability of each event class is known, we can find the probability of any outcome. Let A (success) denote one class of events; let B (failure) denote the other class. Further, let p represent the probability of event A (i.e., the proportion of the sample space occupied by A

⁸Note that $N! = N(N-1)(N-2) \dots [(N-r)+1](N-r)[(N-r)-1] \dots (1)$. Write out $N!$ in the numerator and $(N-r)!$ in the denominator and cancel the terms common to both numerator and denominator and express $(N-r)[(N-r)-1] \dots (1)$ as $(N-r)!$ to obtain Equation 9.11.

events):⁹ $q = 1 - p$ denotes the probability of event B . The probability of a particular permutation, for example, four A 's (successes) followed by a B (failure), then another A (success), (A, A, A, A, B, A), is $p \cdot p \cdot p \cdot p \cdot q \cdot p = p^5q$.

Notice that the result, $p^r q$, is the same no matter where in the sequence the outcome B falls. Thus, the probability of any sequence of N independent Bernoulli events depends only on the probability of event A on any trial (p), and the number of A 's in the sequence (r). In other words, the probability of a given sequence of r successes in N events, where p is the probability of outcome A on any of the N independent trials and $q = 1 - p$ is the probability of outcome B (a non- A event on any trial), is:

$$p^r q^{N-r} \quad (9.13)$$

When tossing a pair of fair die, what is the probability of (7, 7, 7, not-seven, 7) in five tosses? In Table 9.1, we see that in the sample space, the proportion of 7's is $6/36 = 1/6 = p$. The probability of a non-seven then is $q = 1 - p = 1 - 1/6 = 5/6$. Thus, from Equation 9.13, the probability of this sequence is:

$$p^4 q^{5-4} = p^4 q = (1/6)^4 (5/6) = 5/7,776 = .000643$$

or roughly 6 chances in 10,000.

Rarely would our interest be in a particular sequence or permutation; we would usually be more interested in the probability of obtaining four 7's in five tosses, for example. Note that there are five different permutations, each with the same independent probability (.000643), that would result in the same combination; consequently, the probability of one of the five occurring is $5(.000643) = .003215$, or about 3 chances in 1,000.

More generally, from Equation 9.12, note that there are $N!/[(N-r)!]$ sequences that result in r successes in N trials. If $N = 5$, and $r = 4$, there are five sequences that result in this combination.

$$\frac{N!}{r!(N-r)!} = \frac{(5 \cdot 4 \cdot 3 \cdot 2 \cdot 1)}{(4 \cdot 3 \cdot 2 \cdot 1)(1)} = 5$$

Generalizing the above rationale, we come to the following conclusion. Where the probability of the result A is p and the probability of a non- A result is $1 - p = q$, the probability of observing result A in r of N independent occasions is:

$$\frac{N!}{r!(N-r)!} (p^r q^{N-r}) \quad (9.14)$$

Keep in mind that $p^r q^{N-r}$ is the probability of any one of the $N!/[(N-r)!]$ sequences that result in r successes in N trials. Suppose there is a ten-item multiple choice test, with each item containing four options. What is the probability of obtaining a score of 80% from

⁹To be completely consistent in notation, since a parameter is denoted, π rather than p should be used. We shall compromise in this application, however, and use the conventional symbols, p and q .

random guessing? In this situation, $N = 10$, $r = 8$, and the probability of a success on any item is $p = 1/4$. The probability of guessing correctly on 8 (not 8 or more) of the 10 items is:

$$\frac{10!}{8!(10-8)!} \left(\frac{1}{4}\right)^8 \left(\frac{3}{4}\right)^2 = 45 \left(\frac{1}{65,536}\right) \left(\frac{9}{16}\right) = \left(\frac{405}{1,048,576}\right) = .000386$$

or less than 4 chances in 10,000.

What is the probability of answering 8 or more of the 10 items correctly? We need to find the probability of answering 9 items correctly, and the probability of answering all 10 questions correctly, and sum the three probabilities. The probability of 9 from Equation 9.14 is $(10!/9!)[(1/4)^9(3/4)] = 10(1/262,144)(3/4) = 30/1,048,576$; the probability of answering all ten correctly is $(1/4)^{10} = 1/1,048,576$. Thus the probability of answering 8 or more of the 10 four-option questions correctly merely by guessing is only $(405 + 30 + 1)/1,048,576 = 436/1,048,576 = .0004158$.¹⁰

The *binomial expansion* is an application of the binomial theorem you were probably introduced to in high school algebra. It is a general expression that gives the probabilities for any number of outcomes of an event A in N independent Bernoulli trials, where p is the probability of event A on any one trial. The number of A 's is the exponent for p in any of the $N + 1$ terms of the expansion:

$$(p + q)^N = p^N + \frac{N!}{(N-1)!1!} p^{N-1}q + \frac{N!}{(N-2)!2!} p^{N-2}q^2 + \dots + \frac{N!}{1!(N-1)!} pq^{N-1} + q^N \quad (9.15)$$

The binomial expansion or distribution can be used to compare any dichotomous set of observations with a theoretical distribution to answer such questions as,

"In the distribution of male offspring to mothers who carry the gene for hemophilia, what is expected if hemophilia is a Mendelian recessive trait?"

"Are more babies male than female?"

"Do husbands score higher on need-for-achievement measures than do their wives?"

"In a mental telepathy experiment, is the number of 'hits' greater than chance?"

"Can you toss more than 50% heads in a series of coin tosses?"

"Can a naive examinee beat the odds and obtain a score that is greater than can be expected from chance alone?"

"When faced with true-false questions, are examinees more likely to guess T than F?"

"Is the proportion of dropouts who are male greater than the proportion who are female?"

9.11 THE BINOMIAL AND SIGN TEST

The *sign* test is a special "non-parametric"¹¹ application of the binomial distribution in which there are N paired observations, such as matched pairs in an experiment (E vs. C). If

¹⁰Or 4.16×10^{-4} in scientific notation (often appearing as 4.16E-4 in computer output).

¹¹"Non-parametric" is an unfortunate term applied to statistical methods that apply to data that represent nominal or ordinal scales of measurement, or to a data set where normality and other assumptions are not made about parameters of the distribution of observations in the parent population (Conover, 1980, p. 92). "Distribution-free" is a synonym of "non-parametric."

the treatment has no effect, we should expect the E pair-member to surpass the C pair-member in one-half ($p = .5$) of the N comparisons. Suppose only 5 matched pairs are available. We randomly assign one pair-member to get the E treatment, the other pair-member serving as the control. On the posttest, we find that in 4 of the 5 comparisons E outperformed C . How probable is it that we would observe that E outperformed C in 4 or more of the comparisons by chance? Based on Equation 9.15, the probability of exactly 4 "hits" is $5(1/2)^4(1/2) = 5(1/2)^5 = 5/32$; the probability of 5 "hits" is $(1/2)^5 = 1/32$; thus, the probability of 4 or more "hits" in 5 comparisons is $5/32 + 1/32 = 6/32 = .1875$, or roughly 1 chance in 5. Our expectation that E is effective lacks strong evidence. However, if we continue to add cases with the same success ratio, our case will become much stronger.¹²

In the sign test, and many other applications of the *binomial test*, when $p = q = .5$, the product of each of the $N + 1$ " pq " terms equals p^N , which greatly simplifies the computation. If a basketball player makes 50% of his free throws, and if he takes 12 free throws in a game, how likely is it that he will make 10 or more? Since $p^N = (1/2)^{12} = 1/4,096$; the probability of making all 12 is $1/4,096$; the probability of making 11 is $12/4,096$; the probability of making 10 is $[(12)(11)/2]/4,096 = 66/4,096$. Thus, the probability of making 10 or more of the 12 free throws is $(1+12+66)/4,096 = 79/4,096 \approx .02$ (not very likely). Note, however, the probability of missing 10 or more of the 12 has the same probability.¹³

Table O in the Appendix gives the probability for observing r A 's in N events when $p = .5$. Suppose we want to investigate whether this generation is taller than the previous generation, as we have been led to believe. We ask the twenty-five members of the statistics class to compare their heights with that of their parents of the same sex at the same age. If this generation is no taller than the previous generation, we would expect .5 = 50% of the students to be taller than their parents. We collect data and find that twenty of the twenty-five are taller than their same-sex parent. How probable is it that we would observe 20 A 's out of 25 = N Bernoulli trials? In Table O, find the row for $N = 25$ and the column for $N - 5 = 25 - 5 = 20$ A 's (or for 5 B 's). The intersection contains the value .00204, which is the probability that twenty or more A 's would be observed in twenty-five independent events when $p = .5$. Since our result is so unlikely if this generation is no taller than the previous generation, we conclude that this generation is taller.¹⁴

9.12 INTUITION AND PROBABILITY

The study of probability can be interesting and entertaining. Historically, the concepts of probability evolved in connection with games of chance. Those who make use of probability theory are generally awed by the intricacy and excitement of the system and the way in which it produces results that are often quite in disagreement with intuition (unless one's intuition has been developed by experience with calculated probabilities). A few examples of surprising results will illustrate the untrustworthiness of intuition.

First, the classic "birthday problem": What is the probability that at least two people in

¹²Can you confirm that the probability of 8 success in 10 trials is $(45 + 10 + 1)/1,024 = 56/1,024 = .055$?

¹³This problem makes the dubious assumption that the probabilities are independent. If there is any validity to the notion that on a given occasion a player may get "hot," the model does not hold. Whether there are nonrandom sequences such as hitting streaks or slumps in baseball (or whether they are just unusual sequences that are expected infrequently) can be tested by the "runs" test (Dixon & Massey, 1983).

¹⁴In Chapter 11, you will see that we have just rejected the statistical hypothesis that the parameter π is .5.

a group of twenty-three have the same birthday anniversary? Assume that the people are drawn randomly from a population of persons in which all 365 birthdays (not counting 29 February) are equally likely. One often obtains intuitive guesses that the probability is .10 or .001 or even smaller. Surprisingly, it is more likely than not that the probability that at least two people out of twenty-three have the same birthday is .507! Naturally the probability is even higher as the size of the group increases; it is practically certain that in a group of 150 persons at least two people will have the same birthday (Feller, 1957, pp. 31–32.).

A certain TV show has three “windows.” Behind one there is an expensive prize; there is nothing of real value behind the other two. The contestant keeps what is behind the window chosen. Suppose that, after the contestant selects a window, the host always raises one of the two other windows—behind which there is nothing of value, then gives the contestant the choice to remain with the initial choice or to change to the other unopened window. What is the wiser choice? You may feel that the odds started out even, and remain even, so that there is no advantage or disadvantage to remaining with the original choice. If your intuition is that your odds of winning are increased by changing, you are right. The odds of winning change from $1/3$ to $2/3$ when the contestant leaves the initial choice.

What is the probability that a student will obtain a score of 75 on a 100 item true-false test by guessing randomly on each question, given that the average¹⁵ chance score, μ , is 50 and the standard deviation, σ , is 5? The score of 75 is $z = 5$ standard deviations above the chance mean. From Table A in the Appendix, the probability can be read as less than one chance in 1,000,000! What is the probability of a score above 60%? Since 60% is two standard deviations above the mean, only one randomly guessing student out of fifty will obtain a score greater than 60.

The “gambler’s fallacy” represents another example in which intuitive notions of probability often lead to erroneous conclusions. If the football captain has won the coin flip by calling heads for the first three games, should he change to tails on the next flip? If a craps shooter failed to throw a 7 on ten straight throws of a pair of dice, is he more likely to throw a 7 on the next throw than if he has thrown three 7’s in a row? If the first four children in a family are boys, are the chances that the next child will be a girl different from what they would be if the four previous children had been two boys and two girls or all girls? If you think so, shame on you—you are guilty of the “gambler’s fallacy.” If the probabilities of the event in question are independent, as they are in the examples above, the probability of a future event is unaffected by any pattern of past results. Whatever the number of heads, prior to a given toss, the probability of heads on the next toss of a fair coin is .5. This is confirmed by the conditional probability equation (Equation 9.7). In the sample space of four tosses of a coin there are sixteen events or permutations: *HHHH, HHHT, HHTH, HHTT, ... TTTT, TTTT*. If *A* is *HHH* for the first three tosses, then $P(A) = 2/16 = 1/8$; and if *B* is *H* for the fourth toss, then $P(A \cap B) = 1/16$, the probability of head (*B*) given three previous heads (*A*), that is:

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{1/16}{1/8} = .5$$

¹⁵ $\mu = k/a$, where k is the number of items on the test and a is the number of response options per item. The standard deviation of the chance score is $\sigma = \sqrt{k\pi(1-\pi)}$, where $\pi = 1/a$ (see Hopkins, Stanley, & Hopkins, 1990). For large k , the distribution will be approximately normal.

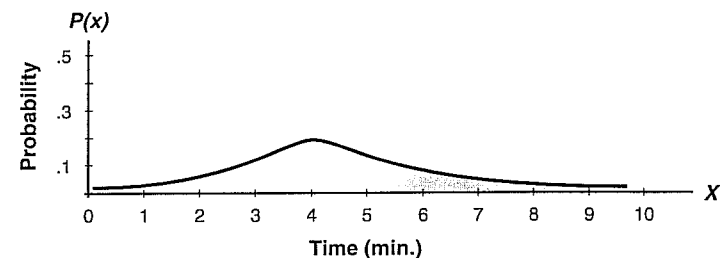


FIGURE 9.3 The Probability Density Function of the Variable *X*, Time Required to Solve a Puzzle

9.13 PROBABILITY AS AN AREA

The probabilities of observing values of continuous variables, for example, height, are conveniently represented by mathematical curves known as probability distributions. Suppose a continuous random variable *X* takes on values from 0 to 10. For example, *X* could be the time required for students to solve a certain puzzle. A student may solve it almost immediately, or she may take as long as ten minutes. Presumably, the length of time required to solve the problem is known for a large number of different subjects. Figure 9.3 represents a graph drawn in which the “time to solution” is graphed against the “proportion of subjects requiring that time.”

The proportion of subjects requiring between two and four minutes to solve the puzzle can be regarded as the probability that a subject selected at random from the population will require that amount of time to solve the puzzle. The entire area under the curve in Figure 9.3 is set to 1, so the area under the curve between any two points X_1 and X_2 is the probability that a randomly selected subject will require between X_1 and X_2 minutes to solve the puzzle. The probability that a randomly selected subject will take more than 5.3 minutes is equal to the shaded area in Figure 9.3. What area corresponds to the probability that a randomly selected subject will take less than 1 minute? (The small area between 0 and 1.)¹⁶ If the proportion of the area under the curve in Figure 9.3 between 6 and 10 were .17, then in a group of 100 randomly chosen subjects we would expect about seventeen persons to take between six and ten minutes to solve the puzzle.

The statistician frequently plots the values a continuous random variable can assume so that the area between any two values of the variable equals the probability that the variable will assume a value between those two values. The resulting graph is called a probability density function. The graph can often be expressed as a mathematical function so that the ordinate $P(X)$ can be found by substituting any value of the random variable *X*. For ex-

¹⁶Theoretically, the probability that a person will take *exactly* one minute (or any other *precise* value), for example, 1.000 ... minutes, is zero; in such situations, the height of the curve is the “probability density” of the value.

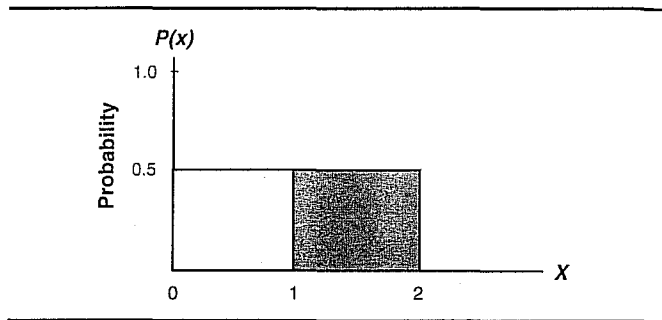


FIGURE 9.4

Probability Density Function of the Variable X That Assumes All Possible Values between 0 and 2 with Equal Probability

ample, assume X is a random variable that can take on any value between 0 and 2 with equal probability. If $P(X) = 1/2$ for all X , then the resulting graph (Figure 9.4) will be the probability density function of X . For example, the area under the curve of the rectangle in Figure 9.4 is exactly 1 (i.e., 0.5×2.0). The lightly shaded area is the probability that X takes on a value between 0 and 1; the probability equals .5.

9.14 COMBINING PROBABILITIES

Suppose we perform a sign test to investigate our research hypothesis that husbands express a higher level of need-for-achievement than do their wives, and suppose the study was replicated by two other researchers. Suppose all three studies found a slight difference in favor of the husbands with probability levels of .09, .06, and .11. None of these studies separately would give convincing evidence of the research hypothesis, but does your intuition suggest that a combination of the three studies would lead to a different conclusion?

There are several methods for addressing this situation, but the "most serviceable under the largest range of conditions is the method of adding z 's" (Rosenthal, 1978). The "method of adding z 's," also known as the Stouffer method, is quite simple and direct: First, convert each of the p -values to its corresponding z -score from Table A. Second, sum the z 's and divide by the square root of the number of probabilities being combined, and third, find the p -value of the obtained z -value from Table A.

For our three p -values of .09, .06, and .11, the corresponding z -scores are 1.341, 1.555, and approximately¹⁷ 1.225; their sum is 4.121, which, divided by the square root of three (1.732), is 2.379. Note from Table A, the proportion of the area above $z = 2.38$ in a normal distribution is .0087; thus, the probability of obtaining three independent p -values of .06, .09, and .11 is less than .01. Hence, taken as a composite set of information, the conclusion is warranted that husbands do indeed have a higher level of need-for-achievement than do wives. Note however, the p 's must be independent; the procedure cannot be used within a single study having multiple measures.

¹⁷This is best done using a spreadsheet (e.g., using the NORMDIST function in EXCEL), although the results from Appendix Table A are quite acceptable.

9.15 EXPECTATIONS AND MOMENTS

Moments are characteristics of distributions defined in terms of expectations. The definition of the expectation of a random variable will be considered first.

Definition: If X is a discrete random variable that takes on the values X_1, X_2, \dots, X_N with probabilities p_1, p_2, \dots, p_N , then the *expectation* of X denoted by $E(X)$ is defined as:

$$E(X) = p_1X_1 + p_2X_2 + \dots + p_NX_N = \sum_j^N p_jX_j \quad (9.16)$$

where $p_1 + p_2 + \dots + p_N = 1 = \sum_j p_j$.

Another symbol denoting the expectation of X is μ , the Greek lowercase letter mu. $E(X) = \mu$, the mean of the population of X 's.

The names "expectation" and "expected value" are synonymous. Some examples of expectations are as follows:

1. Suppose X is the random variable that has six possible values, 1, 2, ..., 6. The events of the sample space could be the six sides of a die. Assume that a probability of 1/6 is associated with each value of X . What is the value of $E(X)$? From Equation 9.16:

$$\begin{aligned} E(X) &= \sum_j^6 p_jX_j = \frac{1}{6}(1) + \frac{1}{6}(2) + \frac{1}{6}(3) + \frac{1}{6}(4) + \frac{1}{6}(5) + \frac{1}{6}(6) = \frac{1}{6}(1+2+\dots+6) \\ &= \frac{21}{6} = 3.5 \end{aligned}$$

In this example, $E(X) = \mu = 21/6 = 3.5$. In repeatedly rolling the die, one can expect to average 3.5 points.

2. A particular slot machine has payoffs of \$0.00, \$0.50, \$1, \$2, and \$25. The probabilities associated with each of these occurrences are .80, .15, .04, .01, and .001, respectively. Define a random variable X that takes on the four values 0, 50, 100, 200, and 2500 cents with probabilities .80, .15, .04, .01, and .001. What is the value of $E(X)$?

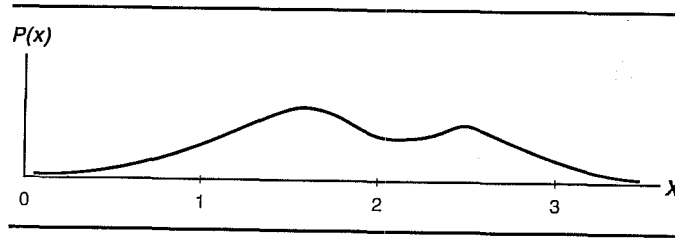
$$\begin{aligned} \mu &= E(X) = .80(0) + .15(50) + .04(100) + .01(200) + .001(2500) \\ &= 0 + 7.5 + 4.0 + 2.0 + 2.5 = 16 \end{aligned}$$

If it costs 25 cents for each trial on this slot machine, would you like to play? If you feed \$25 into the machine, how much can you expect to lose? (\$9)

3. Let X be the four random values that correspond to the number of heads in four flips of a fair coin. X can take on the values 0, 1, 2, 3, and 4. Find $E(X)$. If you write out the sixteen equally probable events in the sample space, that is, $HHHH, HHHT, \dots, TTTT$, you will find the probabilities associated with 0, 1, 2, 3, 4 are 1/16, 1/4, 3/8, 1/4, and 1/16, respectively. Thus:

$$E(X) = \sum_j^4 p_jX_j = \frac{1}{16}(0) + \frac{1}{4}(1) + \frac{3}{8}(2) + \frac{1}{4}(3) + \frac{1}{16}(4) = 2$$

FIGURE 9.5
Probability Distribution
of X



"In the long run," one can expect to average two heads in four random tosses of a fair coin.

If X is a continuous variable instead of a discrete one, then an algebraic function describes the form of its probability distribution. If X is continuous, one cannot assign a probability to a single value of X . Instead, statements about the probability that X lies in an interval are made. For these reasons, the definition given for $E(X)$ in Equation 9.16 cannot be applied to a continuous random variable. Unfortunately for those without recourse to knowledge of integral calculus, it is difficult to define the expectation of a continuous variable.

Suppose X is a continuous random variable and the probability distribution of X looks like the one in Figure 9.5. There is an algebraic rule that gives the height of the curve in Figure 9.5 for every value of X . The area under the curve is 1 unit. The probability that X will assume a value between, for example, 2 and 3 is equal to the area under the curve between those two points.

Definition: The *expectation* of the continuous random variable X is the sum of the products formed by multiplying each value that X can assume by the height of the probability function curve above that value of X .

Since X can take on infinitely many values with continuous random variables, you might wonder how you could multiply each of the separate values of X by the height of the curve at X to find its expectation. This is the problem that recourse to the integral calculus solves. If you are not familiar with calculus, take it on faith that it can be done, in a precise but somewhat indirect way, by "integration."

The expectation of a continuous random variable X is denoted by $E(X)$ or μ , as is the expectation of a discrete variable.

9.16 CHAPTER SUMMARY

The probability of event A can be viewed as the ratio of the number of points of A to the total number in the sample space. Often intuitive notions regarding probability are quite inaccurate.

If events A and B are mutually exclusive, the probability of either A or B , $P(A \cup B)$, is $P(A) + P(B)$. If events A and B are not mutually exclusive, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, where the last term is the probability of both A and B .

If event A has probability $P(A)$, the probability that A will occur n times in n independent trials is $P(A)^n$.

The probability of event B , given event A , is conditional probability, $P(B|A)$, and equals $P(A \cap B)/P(A)$.

Each unique ordering or arrangement of n objects is a permutation. There are $n!$ permutations of n objects. The set of objects irrespective of order is a combination. The number of combinations of n things taken r at a time is $n!/[r!(n-r)!]$.

9.17 SUGGESTED COMPUTER ACTIVITY

From spreadsheet or statistical software, select a random sample of 25 numbers between 1 and 1,000.

MASTERY TEST

1. What is the probability of tossing a "6," $P(6)$, with one toss of a die?
2. How many sample points are there in the sample space in question 1?
3. Are the sample points mutually exclusive?
4. What is the probability of not tossing "6," $P(\bar{6})$?
5. What is the probability of tossing "6's" in two tosses of a die?
6. Given that one "6" has been tossed, what is the conditional probability that the second toss will also be a "6"?
7. On a five-option multiple choice test, what is the probability of selecting the correct answer from a random guess?
8. What is the probability of correctly guessing the right answer on all ten items of a five-option multiple choice test?
9. Although 651412 and 214165 are different permutations, they represent a single _____.
10. What is the probability of correctly guessing ten of ten true-false questions?
11. How many permutations are there in the previous question? (For example, T, T, F, T, T, F, F, F, T, T, is one permutation.)
12. How many different doubles teams in tennis are possible with a class of twenty members?
13. The probability of throwing two consecutive "snake eyes" with the toss of a pair of dice is $(1/36)^2 = 1/1,296$. If you have just tossed "snake eyes," what is the probability the next toss will be "snake eyes"?
14. Which of these random variables are discrete and which are continuous:
 - (a) number of students enrolled in a statistics class
 - (b) running speed of ten-year-olds
 - (c) result from the toss of a pair of dice
 - (d) height of adult males
15. What is the expected number of girls in two-child families if the probabilities associated with 0, 1, and 2 girls are $1/4$, $1/2$, and $1/4$, respectively?
16. Probability density pertains to
 - (a) continuous random variable
 - (b) discrete random variable
17. If 25% of the area in a probability distribution falls between 90 and 100, what is the probability that a case selected at random will fall between 90 and 100?

18. If event A influences the probability of event B , events A and B are not ____.
19. What is the probability of drawing four aces without replacement from a deck of fifty-two cards? (Hint: probability of ace on first card is $4/52$; on second, $3/51$; etc.)

PROBLEMS AND EXERCISES

- Let a pack of fifty-two playing cards be the sample space S of interest. Determine the probabilities of each of the following events:
 - A is "a card is the ace of spades." Find $P(A)$.
 - B is "a card is an ace." Find $P(B)$.
 - C is the event that "a card is a spade." Find $P(C)$.
 - D is "a card is a diamond," and C is "a card is a spade." Find $P(D \cap C)$.
 - C is "a card is a spade" and B is "a card is an ace." Find $P(C \cup B)$.
- Suppose that in a certain locale 3% of the children of kindergarten age have severe perceptual problems and 6% of the children of the same age have emotional problems. Also, 1.5% of the same group of children have both perceptual and emotional problems. Children suffering from either problem or both must receive teaching apart from normal pupils. What is the probability that a child entering kindergarten will require special teaching, that is, will have either perceptual or emotional problems or both?
- (a) Find $15!/13!$
(b) Find $6!/3!(6-3)!$
(c) What is the value of N if $(N+1)!$ is exactly ten times larger than $N!$?
- An experimenter wishes to have subjects learn a list of paired associations in all possible orders of the six pairs. Each subject can learn the list only once. How many subjects would be required if a different subject is required for every possible ordering of the six pairs?
- (a) How many combinations are there of six things taken four at a time—find $\binom{6}{4}$.
(b) Find $\binom{6}{1}$, $\binom{6}{2}$, $\binom{6}{3}$, $\binom{6}{4}$, and $\binom{6}{5}$.
- The varsity basketball team has twelve members. How many possible "starting fives"—the five players who start the game—could the coach form from his team of twelve players?
- (a) Verify that $\binom{6}{0} + \binom{6}{1} + \binom{6}{2} + \binom{6}{3} + \binom{6}{4} + \binom{6}{5} + \binom{6}{6}$ is equal to 2^6 . (In general, $\sum_{i=0}^N \binom{N}{i} = 2^N$)
(b) How many different combinations are there of six things, in groups of size 0 to 6 inclusively, that is, what is $\sum_{i=0}^6 \binom{6}{i}$?
- How many five-item tests can be formed by ten items split into two tests of five items each?
- A student takes a ten-item true-false test but does not know the answers to five of the items. If he guesses randomly between true and false on each of the five items,
 - What is the probability that he will earn a perfect score of 10?
 - What is the probability that he will guess incorrectly on all five questions?
 - What is the probability that his score will be 6, 7, 8, or 9? (Hint: Subtract probabilities for scores of 5 and 10 from 1.00.)
- In a fictitious experiment, convicts volunteered for study of the causal relationship between smoking and lung cancer. The convicts were matched into five matched pairs so that both pair mates were of the same age. Within each pair of convicts, a coin was flipped to determine which convict would continue smoking two packs of cigarettes a day and which one would not be allowed cigarettes for the duration of the experiment. At the end of the ten-year experimental period, the five smokers in each pair had lung cancer; none of the nonsmokers had lung cancer. Suppose that at the outset of the experiment, a convict in each pair had undiagnosed lung cancer.

- What is the probability that the five initially cancerous convicts would be randomly assigned to be the five experimental smokers?
 - If there had been ten pairs, rather than five, what is the probability that the cancerous convict in each of the ten pairs would have been assigned to the smoking group?
- In the general population, Stanford-Binet IQ's are nearly normally distributed with a mean of 100 and a standard deviation of 16. By referring to Table A in the Appendix, determine the following probabilities:
 - A randomly sampled person will have an IQ between 80 and 120.
 - A randomly sampled person will have an IQ above 140.
 - Three independently randomly sampled persons will all have IQ's above 92.
 - The variable X takes on the values 0, 1, 2, 3, and 4 with probabilities 0, $2/5$, $1/5$, $1/5$, and $1/5$, respectively. What is the value of $E(X)$, the expected value of X ?
 - The sample space for tossing a pair of dice is given in Table 9.1.
 - Determine the probability for each value of X , for 2, 3, 4, ..., 12.
 - What is the expected value of X , $E(X)$?
 - What is probability that $X \neq 7$?
 - What is the probability of "7" on three consecutive throws?
 - Given the consecutive "7's," what is the probability of "7" on the next toss?
 - What is the probability of "7's" on four of five tosses?
 - What is the probability of "7's" on four or more of five tosses?
 - In an experiment comparing a new product with the standard product, the new product was preferred by 23 of the 30 subjects in the taste test. Is the new product better, or can this be explained by chance? Use Table O to learn the probability of 23 A's out of 30 Bernoulli events.

ANSWERS TO MASTERY TEST

- $1/6$
- 6
- yes
- $5/6$
- $(1/6)^2 = 1/36$
- $P(B|A) = (1/36)/(1/6) = 1/6$
- $1/5$
- $(1/5)^{10} = .0000001$
- combination
- $(1/2)^{10} = 1/1,024$
- 1,024
- $\binom{20}{2} = \frac{20!}{2!(20-2)!} = 190$
- $1/36$
- (a) and (c) are discrete, (b) and (d) are continuous.
- $E(X) = 1/4(0) + 1/2(1) + 1/4(2) = 1$
- (a)
- .25
- independent
- $(4/52)(3/51)(2/50)(1/49) = 24/6,497,400 = .0000037$

ANSWERS TO PROBLEMS AND EXERCISES

- (a) $P(A) = 1/52$
(b) $P(B) = 4/52 = 1/13$
(c) $P(C) = 13/52 = 1/4$
(d) $P(D \cap C) = 0/52 = 0$
(e) $P(C \cup B) = 13/52 + 4/52 - 1/52 = 16/52 = 4/13$
- A: a child has perceptual problems.
B: a child has emotional problems.
 $A \cap B$: a child has both perceptual and emotional problems.
 $P(A) = .03$, $P(B) = .06$, $P(A \cap B) = .015$.
 $P(A \cup B) = .03 + .06 - .015 = .075$

3. (a) 210
 (b) 20
 (c) If $N = 9$, then $(n + 1)! = 10! = 10(9!)$.
4. $6! = 720$
5. (a) 15
 (b) 5, 10, 10, 5, and 1
6. $\binom{12}{5} = 792$
7. (a) $1 + 4 + 6 + 4 + 1 = 16$; $2^4 = 16$
 (b) $2^6 = 64$
8. $\binom{10}{5} = 252$
9. (a) $P(A)^n = (1/2)^5 = 1/32$ or .03125
 (b) $1/32$ or .03125
 (c) $1 - 1/32 - 1/32 = 30/32 = 15/16$ or .9375
10. (a) $(1/2)^5 = 1/32$
 (b) $(1/2)^{10} = 1/1,024$
11. (a) .7888
 (b) .0062
- (c) $(.6915)^3 = .3307$
12. $E(X) = 2\frac{1}{2}$
13. (a) $1/36, 1/18, 1/12, 1/9, 5/36, 1/6, 5/36, 1/9, 1/12, 1/18, 1/36$
 (b) $E(X) = \sum_j p_j X_j = 7$
 (c) $5/6$
 (d) $(1/6)^3 = 1/216$
 (e) $1/6$
- (f) $\frac{N!}{(N-1)!!} p^{N-1} q, \left(\frac{5!}{4!}\right) = 5$, hence
 $5 \left(\frac{1}{6}\right)^4 \left(\frac{5}{6}\right) = \frac{25}{7,776} = .0032150$
- (g) $p^N = (1/6)^5 = 1/7,776$; thus, $26/7,776 = .0033436$
14. The probability of observing 23 or more A's when $N = 30$ is only .00261, good news for the marketing department.

10

STATISTICAL INFERENCE: SAMPLING AND INTERVAL ESTIMATION

10.1 OVERVIEW

In the preceding chapters, statistical inference has been of incidental concern—only a minor theme. Beginning with this chapter, we will focus on estimating parameters using inferential statistical methods. One of the primary purposes of statistical methods is to allow generalizations about populations using data from samples. This chapter introduces ideas that are of fundamental importance in all succeeding chapters.

Nearly all public opinion polls and surveys, such as the Gallup and Harris polls, involve selecting a sample, obtaining data on that sample, and then making inferences about the entire population. Rarely are all members of the population observed; usually only a small fraction of the elements in the population is sampled. The Nielsen ratings of the popularity of television programs are based on the viewing habits of a sample of less than one home in 10,000 (.01%) in the population. The computerized projections of winners in political elections are nothing more than sophisticated applications of the concepts of this chapter. Before considering the theory underlying statistical inference, some fundamental definitions and concepts must be reviewed.

10.2 POPULATIONS AND SAMPLES: PARAMETERS AND STATISTICS

The principal use of statistical inference in empirical research is to obtain knowledge about a large class of persons, or other statistical units, from a relatively small number of persons. Inferential statistical methods employ *inductive reasoning*—reasoning from the particular to the general and from the observed to the unobserved. Inferential statistical reasoning addresses such questions as: “What can I say about the age at which the average child in the United States (the population) first utters a sentence, given that the average was 202 weeks for a sample of twenty-five children?” Any exhaustive (finite or infinite) set or collection of things (units) that we wish to study, or about which we wish to make inferences, is called a