3. Let $D = \{(x_1, x_2, x_3): x_1 = 0, 1;\ x_2 = 0, 1;\ x_3 = 0, 1\}$ and define

$$h[(x_1, x_2, x_3)] = x_1 + x_2 - x_3.$$

What is the range of h?

4. Define $C = \{(x_1, x_2): x_1 = 1, 2, 3, 4, 5, 6;\ x_2 = 1, 2, 3, 4, 5, 6\}$ and let

$$g[(x_1, x_2)] = x_1 + x_2, \qquad \text{for } (x_1, x_2) \in C.$$

Specify

$$A_7 = \{(x_1, x_2): g[(x_1, x_2)] = 7\}$$
$$A_3 = \{(x_1, x_2): g[(x_1, x_2)] = 3\}$$
$$A_{10} = \{(x_1, x_2): g[x_1, x_2)] = 10\}.$$

5. Let $\mathcal{F}$ be the class of all subsets of $S = \{1, 2, 3, \ldots, k\}$. Define $n(A)$ = number of elements in A for $A \in \mathcal{F}$. What is the range of n?

6. Let $\mathcal{K}$ be the class of all subsets of $S = \{1, 2, 3, \ldots, k\}$. Define $p(A) = 1/k$ times the number of elements in A for $A \in \mathcal{K}$. What is the range of p?

7. Let S and n be as defined in problem 5. Show that

$$n(A) \leq n(B), \qquad \text{if } A \subset B,$$
$$n(A \cup B) = n(A) + n(B), \qquad \text{if } A \cap B = \varnothing.$$

2

Probability

We are now ready to begin our study of probability theory itself. The earliest known applications of probability theory occurred in the seventeenth century. A French nobleman of that time was interested in several games then played at Monte Carlo; he tried unsuccessfully to describe mathematically the relative proportion of the time that certain bets would be won. He was acquainted with two of the best mathematicians of the day, Pascal and Fermat, and mentioned his difficulties to them. This began a famous exchange of letters between the two mathematicians concerning the correct application of mathematics to the measurement of relative frequencies of occurrences in simple gambling games. Historians generally agree that this exchange of letters was the beginning of probability theory as we now know it.

For many years a simple relative frequency definition of probability was all that was known and was all that many felt was necessary. This definition proceeds roughly as follows. Suppose that a chance experiment is to be performed (some operation whose outcome cannot be predicted in advance); thus there are several possible outcomes which can occur when the experiment is performed. If an event A occurs with m of these outcomes, then the probability of A occurring is the ratio m/n where n is the total number of outcomes possible. Thus, if the experiment consists of one roll of a fair die and A is the occurrence of an even number, the probability of A is $\frac{3}{6}$.

There are many problems for which this definition is appropriate, but such a heuristic approach is not conducive to a mathematical treatment of the theory of probability. The mathematical advances in probability theory were relatively limited and difficult to establish on a firm basis until the Russian mathematician A. N. Kolmogorov gave a simple set of three axioms or rules which probabilities are assumed to obey. Since the establishment of this firm axiomatic basis, there have been great strides made in the theory of probability and in the number of practical problems to which it is applied.

In this chapter we shall see what these three axioms are and why we might reasonably adopt these rules for probabilities to follow. The axioms do not give any unique value which the probability of an event must equal; rather they express internal rules which ensure consistency in our arbitrary assignment of probabilities. The relative frequency definition of probability which was already mentioned is only one arbitrary way in which probabilities can be computed; it is discussed more fully in Section 2.3 and, as is shown there, does satisfy the axioms. This definition has built within it certain assumptions about the outcomes of the experiment which are not always appropriate. Thus, this arbitrary way of assigning probabilities is not always applicable.

It might aid the reader to discuss briefly at this time the notion of a probability model. An experiment is a physical operation which, in the real world, can result in one of many possible outcomes. For example, if we roll a pair of dice one time the two numbers we might observe can range anywhere from a pair of ones to a pair of sixes. Or, if a particular individual is going to run one hundred yards as fast as he can, the elapsed time from when he starts until when he finishes might be anywhere from nine seconds to thirty seconds. Or, if a particular person lives in a given way in terms of rest and habits of various kinds, his total life span might lie anywhere from zero years to one hundred years. In each of these cases the particular outcome we might observe cannot be predicted in advance, but the total collection of outcomes can be assumed to be known (one of which will be observed when the experiment is completed). When building a probability model for an experiment, we are concerned with specifying: (1) what the total collection of outcomes could be and (2) the relative frequency of occurrence of these outcomes, based on an analysis of the experiment. We are in many senses idealizing the physical situation by restricting the set of possible outcomes; but, to the extent that this idealization does not affect the relative frequency of events of interest, we can profitably use the results of our computations as descriptions of the actual physical experiment. The probability model then consists of the assumed collection of possible outcomes and the assigned relative frequencies or probabilities of these outcomes. The axioms are used to assure consistency in this assignment of probabilities.

It is easy here to give also an indication of the subject matter of statistics that we shall study in the latter portions of this book. Probability theory is concerned with the consistency of assumed probabilities of various outcomes

of an experiment and its implications. One problem of statistics, on the other hand, is concerned with whether or not the actual observed outcomes of the physical experiment (in a series of experiments) are consistent with an assumed probability model; another problem of statistics might be that of using a set of actual experimental outcomes to construct a probability model of the physical experiment. Thus in statistics we shall generally be concerned with problems of inference from a set of observed sample outcomes; the theory of probability will prove very useful in measuring the correctness of these inferences.

It is hoped that these brief statements will help orient the reader for our studies of probability and of statistics.

2.1. Sample Space; Events

As mentioned in Chapter 1, set theory provides a language ideally suited for the efficient study of probability theory. After the following two basic definitions, we shall see in what manner set theory can be utilized.

DEFINITION 2.1.1. An *experiment* is any operation whose outcome cannot be predicted with certainty.

DEFINITION 2.1.2. The *sample space* S of an experiment is the set of all possible outcomes for the experiment.

Let us examine some particular examples which utilize these two definitions.

Example 2.1.1. We roll a single die one time (dice is the plural of die). Then the experiment is the roll of the die. A sample space for this experiment could be

$$S = \{1, 2, 3, 4, 5, 6\}$$

where each of the integers 1 through 6 is meant to represent the face having that many spots being uppermost when the die stops rolling.

Example 2.1.2. We select 1 card at random from a standard deck of 52. The experiment is the selection of a card. We might assume that we have numbered the cards (in some specified order) from 1 to 52; then a sample space for this experiment would be

$$S = \{1, 2, 3, \ldots, 52\}$$

since the particular card that we select in performing the experiment must correspond to exactly one of these integers.

Example 2.1.3. We select a person at random from the student body of UCLA. The experiment is the selection of a student from the total student body. A sample space for the experiment would be the set of all students that could be selected; i.e., S would simply be the roster of students at UCLA at the time the selection is made.

Frequently the performance of an experiment naturally gives rise to more than one piece of information which we may want to record. If we observe two pieces of information every time the experiment is performed, we would reasonably want a sample space that is a collection of 2-tuples, the two positions corresponding to the two pieces of information. Or, if we observe three pieces of information, we would want a sample space of 3-tuples. Or, more generally, if we observe r pieces of information, we would want a sample space of r-tuples. In each of the three examples given above, a single piece of information was generated when performing the experiment; thus each of these sample spaces had 1-tuples as elements. The next three examples discuss experiments in which more than one piece of information is derived.

Example 2.1.4. Suppose our experiment consists of one roll of two dice, one red and the other green. A reasonable sample space for the experiment would be the collection of all possible 2-tuples (x_1, x_2) that could occur where the number in the first position of any 2-tuple corresponds to the number on the red die and the number in the second position corresponds to the number on the green die. Thus, we might use as our sample space

$$S = \{(x_1, x_2): x_1 = 1, 2, 3, \ldots, 6;\ x_2 = 1, 2, 3, \ldots, 6\}.$$

Example 2.1.5. Doug, Joe, and Hugh match coins. The experiment they perform is one flip of three coins. A reasonable sample space for this experiment is the set of 3-tuples, each of which has H (head) or T (tail) in every position. The first position in the 3-tuple corresponds to the face on Doug's coin, the second position to the face on Joe's coin, and the third position to the face on Hugh's coin. This sample space S can be written

$$S = \{(x_1, x_2, x_3): x_1 = H \text{ or } T, x_2 = H \text{ or } T, x_3 = H \text{ or } T\}.$$

Example 2.1.6. We select 10 students, at random, from the student body of the University of Chicago. We would normally restrict the sampling method so that the same student does not occur more than once in the sample (this is called sampling without replacement; see Chapter 6 for a more complete discussion of sampling and its relation to probability). The experiment then is the selection of 10 students. A sample space for the experiment would be the total collection of 10-tuples that could occur, the first element in any 10-tuple corresponding to the first student selected, the second element corresponding to the second student selected, the third element to the third student, etc. For convenience, let us assume that the student body at the University of Chicago consists of 30,000 students and that they have been numbered from 1 to 30,000. Then this sample space may be written

$$S = \{(x_1, x_2, \ldots, x_{10}): x_i = 1, 2, 3, \ldots, 30{,}000;\quad i = 1, 2, \ldots, 10;\ x_i \neq x_j \text{ for all } i \neq j\}.$$

One thing to notice about the sample space of an experiment is that it is not unique; that is, there generally is more than one reasonable way of

specifying all possible outcomes of the experiment. For example, if we roll a pair of dice we could use the set of 2-tuples mentioned in Example 2.1.4 as the sample space, or we could reason that every time we roll two dice the sum of the two numbers that occur could be any of the integers 2 through 12 and adopt this set as the sample space. In Example 2.1.5, Doug, Joe, and Hugh will observe 0 or 1 or 2 or 3 heads when they flip their coins; therefore we could adopt the set having these integers as elements as our sample space rather than the set of 3-tuples mentioned above. It would be equally easy to mention an alternative sample space for Example 2.1.6. As we shall see in the sequel, one sample space is generally easier to use than another for a given experiment. Of course, we shall try to use the one that is easiest for our purposes. Which sample space is used has no effect on the values of probabilities of interest but does affect the ease of computation of such probabilities.

The following two definitions are very basic to the sequel. Misunderstanding of the definition of an event and of when an event has or has not occurred can lead to a great deal of difficulty in many problems.

DEFINITION 2.1.3. An *event* is a subset of the sample space. Every subset is an event.

DEFINITION 2.1.4. An event *occurs* if any one of its elements is the outcome of the experiment.

The sample space used for Example 2.1.1 was

$$S = \{1, 2, 3, 4, 5, 6\}.$$

Then each of the sets

$$A = \{1\}, \quad B = \{1, 3, 5\}, \quad C = \{2, 4, 6\},$$
$$D = \{4, 5, 6\}, \quad E = \{1, 3, 4, 6\}, \quad F = \{2\},$$

is an event (these are not the only events since they are not the only subsets of S). These are all distinct (different) events because no two of these subsets are equal. If we actually were to perform the experiment (roll the die) and we got a 1, then events A, B, and E are said to have occurred since each of these has 1 as an element. Events C and D did not occur since $1 \notin C$ and $1 \notin D$. If we got a 4 when the die was rolled, then we would say that events C, D, and E occurred since 4 is an element of each of them. Notice that no matter which outcome we observe when the experiment is performed, many different events have each occurred (as we shall see, exactly half the possible events occur for any particular outcome).

In Example 2.1.3, which consists of selecting one student at random from the student body of UCLA, the sample space consists of the roster of students at the school. Assuming there are thirty-five thousand students at UCLA, we can represent the sample space by

$$S = \{x: x = 1, 2, 3, \ldots, 35{,}000\}.$$

The sets

$$A_1 = \{1, 50\}, \qquad B = \{2, 76, 140, 64\}, \qquad C = \{10\}, \qquad D = \{x: x = 101, 102, \ldots, 960\},$$

etc., are all events since each is a subset of S. If, when we performed the experiment, we happened to select the student numbered 176, then D occurred since $176 \in D$, but none of the other three events listed occurred. If we happened to select the student numbered 9999, then none of the four events listed occurred since this number doesn't belong to any of them.

We shall frequently want to take a word description of an event and translate it into a subset of the sample space. For example, suppose that in Example 2.1.2 we had numbered the 52 cards in the following way: A, 2, 3, . . . , K of hearts are numbered 1 through 13, respectively; A, 2, 3, . . . , K of diamonds are numbered 14 through 26, respectively; A, 2, 3, . . . , K of spades are numbered 27 through 39, respectively; A, 2, 3, . . . , K of clubs are numbered 40 through 52, respectively. Then we might define the following events in words:

A: a red card is drawn

B: a spade is drawn

C: an ace is drawn

D: a face card is drawn (aces are not counted as face cards).

Then, since an event occurs if any of its elements occurs, we have

$$A = \{1, 2, \ldots, 26\}$$
$$B = \{27, 28, \ldots, 39\}$$
$$C = \{1, 14, 27, 40\}$$
$$D = \{11, 12, 13, 24, 25, 26, 37, 38, 39, 50, 51, 52\}.$$

In Example 2.1.4 a pair of dice is rolled one time. The sample space S is the set of 2-tuples

$$S = \{(x_1, x_2): x_1 = 1, 2, \ldots, 6;\ x_2 = 1, 2, \ldots, 6\}.$$

The events:

A: the sum of the two dice is 3

B: the sum of the two dice is 7

C: the two dice show the same number

are as subsets,

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

Suppose that in Example 2.1.5, Doug, Joe, and Hugh are playing a game called "odd man loses." That is, if two of the coins' faces match and the third person's does not, the third person loses. The sample space is

$$S = \{(x_1, x_2, x_3): x_1 = H \text{ or } T, x_2 = H \text{ or } T, x_3 = H \text{ or } T\},$$

where the first position corresponds to Doug's coin, the second to Joe's, the third to Hugh's. Define the events:

A: Doug loses

B: Doug doesn't lose

C: Joe loses

D: No one loses.

Then, written as subsets, we have

$$A = \{(H, T, T), (T, H, H)\}$$
$$B = \{(H, H, T), (H, T, H), (H, H, H), (T, T, H), (T, H, T), (T, T, T)\}$$
$$C = \{(T, H, T), (H, T, H)\}$$
$$D = \{(H, H, H), (T, T, T)\}.$$

Example 2.1.7. Suppose our experiment consists of a hundred-yard dash involving four college-age sprinters. It is clear that there are many facets of the experiment that we might be interested in such as the name of the winner, the winning time, the order in which the four cross the finish line, the time of the second-place man, or the time of the third-place man. Which facets were of interest would determine the sample space to be used. For example, if we were going to refer only to the winning time we could use

$$S_1 = \{t: 0 \le t \le 15\}$$

(measuring time in seconds). Or, if we were interested in the times of both the first- and second-place men, we could use

$$S_2 = \{(t_1, t_2): 0 \le t_1 \le 15, t_1 < t_2 \le 20\}.$$

In this latter case t_1 is the time of the winner and t_2 is the time of the second-place man; thus the requirement $t_2 > t_1$. For either of these sample spaces we could define the event A: winning time is between 9.45 and 9.65 seconds; using S_1 we would have

$$A = \{t: 9.45 \le t \le 9.65\}.$$

And using S_2 we would have

$$A = \{(t_1, t_2): 9.45 \leq t_1 \leq 9.65, t_1 < t_2 \leq 15\}.$$

Again, then, the sample space is not unique. If A were the only event of interest, we would undoubtedly decide to use S_1 rather than S_2 since this would require keeping track of only the single time of the first-place finisher.

EXERCISE 2.1.

1. Specify a sample space for the experiment which consists of drawing 1 ball from an urn containing 10 balls of which 4 are white and 6 are red. (Assume that the balls are numbered 1 through 10.)

2. Specify a sample space for the experiment which consists of drawing 2 balls with replacement from the urn containing 10 balls (that is, the first ball removed is replaced in the urn before the second is drawn out). Again assume that they are numbered.

3. Specify a sample space for the experiment which consists of drawing 2 balls without replacement from the urn containing 10 balls (that is, the first ball removed is not replaced in the urn before the second is drawn out). Assume that they are numbered.

4. For the sample space given in problem 1, define the events (as subsets):

A: a white ball is drawn
B: a red ball is drawn.

5. For the sample space given in problem 2, define the events (as subsets):

C: the first ball is white
D: the second ball is white
E: both balls are white.

Does $C \cap D = E$?

6. A cigarette company packs 1 of 5 different slips, labelled a, b, c, d, e, respectively, with each pack it produces. Suppose that you buy 2 packs of cigarettes of this brand. What is a good sample space for the experiment whose outcome is the pair of slips you receive with the 2 packs?

7. Suppose that all of the residents of a particular town are bald or have brown hair or have black hair. Furthermore, each resident has blue eyes or brown eyes. We select one resident at random. Give a sample space S for this experiment and define, as subsets, these events:

A: the selected resident is bald
B: the selected resident has blue eyes
C: the selected resident has brown hair and brown eyes.

8. Three girls, Marie, Sandy, and Tina, enter a beauty contest. Prizes are awarded for first and second place. Specify a sample space for the experiment which consists of the choice of the two winners. Define, as subsets, the events:

A: Marie wins
B: Marie gets second prize
C: Tina and Sandy get the prizes.

9. Three cards are selected at random without replacement from a deck which contains 3 red, 3 blue, 3 green, and 3 black cards. Give a sample space for this experiment and define the events:

A: all the selected cards are red
B: 1 card is red, 1 green, and 1 blue
C: 3 different colors occur
D: all 4 colors occur.

10. A small town contains 3 grocery stores (call them 1, 2, 3). Four ladies living in this town each randomly and independently pick a store in which to shop (in this town). Give a sample space for the experiment which consists of the selection of stores by the ladies and define the events:

A: all the ladies choose store 1
B: half the ladies choose store 1 and half choose store 2
C: all the stores are chosen (by at least one lady).

2.2. Probability Axioms

As we shall see in this section, the theory of probability is concerned with consistent ways of assigning numbers to events (subsets of the sample space S) which are called the probabilities of occurrence of these events. It is because we want to have the ability to compute the probability of occurrence of any subset of the sample space that we permit every subset to be called an event. In almost every problem, then, we shall be aware that there are many events which we could define and whose probabilities we could compute in addition to the particular few events that we shall have interest in. It seems much more satisfying to have the ability to compute probabilities for any conceivable event (most of which are not of interest) than suddenly to come across a problem for which we cannot derive an answer because the quantity of interest is not an event.

Most people with an intuitive feeling for what probability should be give a relative frequency interpretation to numbers called probabilities. For example, most would be quick to agree that the probability of a head occurring if we flip a fair coin is one-half, meaning that if the coin is fair then half the time we should observe a head. There are two immediate consequences to a relative frequency motivation for probability. First, the relative frequency of occurrence of something we are sure will occur should be 1; thus its probability should be 1. For example, if we flip a coin one time the relative frequency of the event of observing a head or a tail should be 1, thus the probability of this event should be 1. Second, a relative frequency can never be negative; thus the probability of any event should be nonnegative. These two rules are the first two axioms of an abstract probability theory in order that assignments of probabilities will have these intuitive properties. We shall insist that the probabilities of occurrence of any events satisfy these two requirements.

There is only one other rule or axiom which we shall insist is always satisfied—the additivity property of probability. If we refer to the experiment which consists of one roll of one die, it is only reasonable to expect that the probability that we observe a 1 or a 2 should be equal to the sum of the probability of observing a 1 plus that of observing a 2 since relative frequencies of occurrence have this property. More generally, we would expect the probability of any event A to be equal to the sum of the probabilities of any two nonoverlapping events which together constitute the event A (their union is A). This is the additivity requirement.

Probability can be called a measure applied to the events that can occur when an experiment is performed. To the extent that the assumptions that generate the numerical values of the measure are "correct" for the given problem, the probabilities coincide with the relative frequency notions just discussed. To ensure that this is the case, the probability measures must satisfy the three axioms given below.

Formally, a probability function is a real-valued set function defined on the class of all subsets of the sample space S; the value that is associated with a subset A is denoted by $P(A)$. The assignment of probabilities must satisfy the following three rules (in order that the set function may be called a probability function):

1. $P(S) = 1$
2. $P(A) \geq 0 \qquad$ for all $A \subset S$
3. $P(A_1 \cup A_2 \cup A_3 \cup \cdots) = P(A_1) + P(A_2) + P(A_3) + \cdots$
   if $A_i \cap A_j = \varnothing \qquad$ for all $i \neq j$.

Note that these are the three axioms just discussed. For any experiment, the sample space S plays the role of the universal set; thus any complements which we refer to are taken with respect to S.

Many consequences or theorems can be derived for any probability

function. Let us take a look at some of these now. It will be recalled that $\varnothing \subset S$. Thus our probability measure must assign some number to this event. The number that must always be assigned is 0, as is proved in Theorem 2.2.1.

Theorem 2.2.1. $P(\varnothing) = 0$ for any S.

Proof: $S \cup \varnothing = S$ and thus $P(S \cup \varnothing) = P(S) = 1$ by axiom 1. But $S \cap \varnothing = \varnothing$ so that $P(S \cup \varnothing) = P(S) + P(\varnothing) = 1 + P(\varnothing)$ by axiom 3. Thus $1 + P(\varnothing) = 1$; that is, $P(\varnothing) = 0$. ◀

A second consequence of the assumed axioms is given as Theorem 2.2.2. We shall see many instances in which it saves a great deal of effort in computing probabilities.

Theorem 2.2.2. $P(\bar{A}) = 1 - P(A)$, where $\bar{A}$ is the complement of A with respect to S.

Proof: $A \cup \bar{A} = S$ so $P(A \cup \bar{A}) = P(S) = 1$ by axiom 1. But $A \cap \bar{A} = \varnothing$ and thus $P(A \cup \bar{A}) = P(A) + P(\bar{A})$ by axiom 3. Thus we have established that $P(A) + P(\bar{A}) = 1$, from which the result follows immediately. ◀

Axiom 3 tells us that if two events A and B have no elements in common, then the probability of their union is the sum of their individual probabilities. Theorem 2.2.3 derives a preliminary result which is used to establish Theorem 2.2.4, regarding $P(A \cup B)$ when A and B have elements in common.

Theorem 2.2.3. $P(\bar{A} \cap B) = P(B) - P(A \cap B)$.

Proof: By referring to Figure 2.1, it can be seen that

$$B = (\bar{A} \cap B) \cup (A \cap B).$$

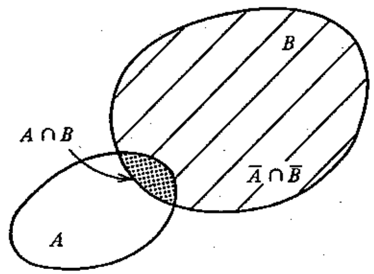

Figure 2.1.

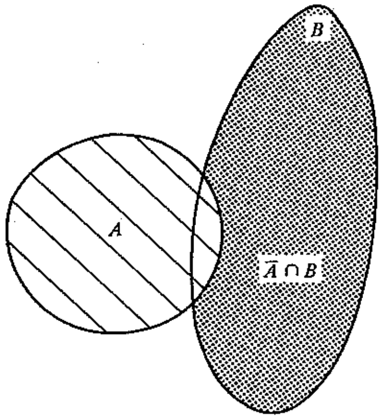

Figure 2.2.

Then $P(B) = P((\bar{A} \cap B) \cup (A \cap B))$. Furthermore, $(\bar{A} \cap B) \cap (A \cap B) = \varnothing$; thus, $P((\bar{A} \cap B) \cup (A \cap B)) = P(\bar{A} \cap B) + P(A \cap B)$, and we have established that

$$P(B) = P(\bar{A} \cap B) + P(A \cap B)$$

from which the desired result follows immediately. ◀

Theorem 2.2.4. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Proof: By referring to Figure 2.2, it can be seen that we can write

$$A \cup B = A \cup (\bar{A} \cap B),$$

and thus $P(A \cup B) = P(A \cup (\bar{A} \cap B))$. Furthermore, $A \cap (\bar{A} \cap B) = \varnothing$ so that

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

by Theorem 2.2.3. Thus

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

Notice that the probability function we have discussed in this section is just a special sort of (real-valued) set function. For every experiment that can be performed we assume that we can define the sample space S which has as elements all possible outcomes; the class of all subsets of S can then be defined and probabilities are just the values assumed by a set function defined

on the class of all subsets. The set function whose values are called probabilities is distinguished by the fact that the three axioms just discussed must be satisfied.

It is important to realize that these axioms will not give a unique assignment of probabilities to events; rather the axioms simply clarify relationships between probabilities that we assign so that we shall be consistent with our intuitive notions of probability. For example, if a rocket has been designed to take a man to the moon, then the experiment which consists of firing the rocket and the man at the moon can be thought of as having two outcomes: success and failure. Success would be the safe arrival of the man on the moon, failure anything else that might occur. Then the axioms do not imply that the probability of the event {success} must be $\frac{1}{2}$ or $\frac{3}{4}$ or .99 or any other particular value. If we denote this probability by p, they do imply that $0 \leq p \leq 1$ and that the probability of the event {failure} must be $1 - p$. Beyond this, p is still unspecified.

Actual specification of the value of p must come from analytical considerations of the experiment performed and the mechanism behind it. For the rocket example just mentioned, this would consist of detailed examination of the rocket design and conditions under which it is to be fired, in addition to any prior test firings or performance data available. Generally, considerations of prior data and their implications regarding the value of p fall into the realm of statistics, the topic for the latter half of this volume.

EXERCISE 2.2.

1. Given $S = \{1, 2, 3\}$, $A = \{1\}$, $B = \{3\}$, $C = \{2\}$, $P(A) = \frac{1}{3}$, $P(B) = \frac{1}{3}$, find:

(a) $P(C)$ (d) $P(\bar{A} \cap \bar{B})$

(b) $P(A \cup B)$ (e) $P(\bar{A} \cup \bar{B})$

(c) $P(\bar{A})$ (f) $P(B \cup C)$.

2. Let S, A, B, C be defined as in problem 1, but now let $P(A) = \frac{1}{2}$, $P(B) = \frac{1}{5}$. Compute the probabilities asked for in (a) through (f).

3. Let S, A, B, C be defined as in problem 1 and let $P(A) = 1$. Compute the probabilities asked for in (a) through (f). Could we let $P(A)$ be 2?

4. Define $S = \{a, b\}$, $B = \{b\}$. Give three different assignments of probabilities to the subsets of S.

5. Prove, from the axioms, that probabilities are monotonic; that is, $P(A) \leq P(B)$ if $A \subset B$.

6. Prove, from the axioms, that $P(A) \leq 1$ for all A.

7. Given an experiment such that $P(A) = \frac{1}{2}$, $P(B) = \frac{1}{2}$, $P(A \cup B) = \frac{2}{3}$, compute:

(a) $P(\bar{A})$ (e) $P(\bar{A} \cup \bar{B})$
(b) $P(\bar{B})$ (f) $P(A \cap \bar{B})$
(c) $P(A \cap B)$ (g) $P(\bar{A} \cap B)$
(d) $P(\bar{A} \cap \bar{B})$ (h) $P(\bar{A} \cup B)$.

8. Given an experiment such that $P(A) = \frac{1}{2}$, $P(B) = \frac{1}{3}$, $P(A \cap B) = \frac{1}{4}$, compute:

(a) $P(A \cup B)$ (d) $P(A \cap \bar{B})$
(b) $P(\bar{A} \cup B)$ (e) $P(\bar{A} \cap \bar{B})$
(c) $P(\bar{A} \cap B)$ (f) $P(\bar{A} \cup \bar{B})$.

9. Is it possible to have an assignment of probabilities such that $P(A) = \frac{1}{2}$, $P(A \cap B) = \frac{1}{3}$, $P(B) = \frac{1}{4}$?

10. If we know that $P(A \cup B) = \frac{3}{8}$ and $P(A \cap B) = \frac{1}{3}$, can we determine $P(A)$ and $P(B)$?

2.3. Single-Element Events and the Equally Likely Case

When we are dealing with an experiment that has a finite number of possible outcomes (thus S is a finite set), the concept of a single-element event becomes rather important; as we shall see, the specification of the values of the probability function P for the single-member events then completely specifies the values of P for all events. First, the definition of a single-element event is as follows.

DEFINITION 2.3.1. A *single-element event* A is a subset of S which has only one element of S belonging to it; that is, if there exists only one $x \in S$ such that $x \in A \subset S$, then A is a single-element event.

For the sample space $\{H, T\}$, the single-element events are $\{H\}$, $\{T\}$. For the sample space $\{1, 2, 3, 4, 5, 6\}$, the single-element events are $\{1\}$, $\{2\}$, $\{3\}$, $\{4\}$, $\{5\}$, $\{6\}$. For the sample space $\{(0, 0), (0, 1), (1, 0), (1, 1)\}$, the single-element events are $\{(0, 0)\}$, $\{(0, 1)\}$, $\{(1, 0)\}$, $\{(1, 1)\}$. As can easily be seen, if S has n elements, then there are exactly n distinct single-element events. The following theorem shows how the probabilities of occurrence of the single-element events imply the probabilities of occurrence of any event, no matter how many elements belong to it.

Theorem 2.3.1. Given a sample space S and any event $A \subset S$,

$$P(A) = P(A_1) + P(A_2) + \cdots + P(A_k)$$

where $A_1, A_2, \ldots, A_k$ are distinct single-element events and $A = A_1 \cup A_2 \cup \cdots \cup A_k$.

Proof: Suppose that $A_1, A_2, \ldots, A_k$ are distinct single-element events. Then $A_i \cap A_j = \varnothing$ for all $i \neq j$ and, by axiom 3, $P(A_1 \cup A_2 \cup \cdots \cup A_k) = \sum_{i=1}^{k} P(A_i)$. Thus, if $A = A_1 \cup A_2 \cup \cdots \cup A_k$, we have

$$P(A) = \sum_{i=1}^{k} P(A_i).$$ ◀

Example 2.3.1. Suppose we roll a six-sided die. Then $S = \{1, 2, 3, 4, 5, 6\}$. The single-element events are $\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}$. Theorem 2.3.1 tells us that if we know the probability of occurrence for each of these six events, we can use this information to compute the probability of occurrence of any other event of interest. For example,

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

In many experiments it is quite reasonable to assume that each single-element event is as likely to occur as any other. For example, if our experiment consists of one flip of a fair coin, it is reasonable to assume that the single-element events of $S = \{H, T\}$ are equally likely to occur. Or, if our experiment consists of one roll of a fair die, it is reasonable to assume that the single-element events of $S = \{1, 2, 3, 4, 5, 6\}$ are equally likely to occur. If we make the assumption that all single-element events of S are equally likely, there is a very simple way in which we can assign probabilities to all the subsets of S and still be consistent with the three axioms.

Assume that we have an experiment with k possible outcomes, and from the description of the experiment we are willing to assume that each of the single-element events is as likely to occur as any other. Then, since the total probability of the whole sample space is to be 1, the common value of the probability for each of the single-element events must be $1/k$. Furthermore, since any event is the union of single-element events (as we just saw in Theorem 2.3.1), the probability of any event $A \subset S$ then is given by the ratio of the number of elements in A (the number of single-element events whose union is A) to the number of elements in S. That is, we use the rule

$$P(A) = \frac{n(A)}{n(S)} \qquad \text{for } A \subset S$$

where $n(A)$ is the number of elements in A (see the last example in Chapter 1, Section 1.4). That this rule will satisfy the three axioms given in Section 2.2 is proved in the next theorem.

Theorem 2.3.2. If S has k elements the rule

$$P(A) = \frac{n(A)}{n(S)}$$

satisfies the three axioms for a probability function.

Proof: If S has k elements, then $n(S) = k$.

$$P(S) = \frac{n(S)}{n(S)} = \frac{k}{k} = 1$$

so axiom 1 is satisfied. If A is any subset of S, it contains a nonnegative number of elements; i.e., $n(A) \geq 0$ for all $A \subset S$. Then

$$\frac{n(A)}{k} = \frac{n(A)}{n(S)} = P(A) \geq 0 \qquad \text{for all } A \subset S$$

and axiom 2 is satisfied. If $A \cap B = \varnothing$, then A and B have no elements in common and we would know that $n(A \cup B) = n(A) + n(B)$. Thus

$$\frac{n(A \cup B)}{n(S)} = \frac{n(A)}{n(S)} + \frac{n(B)}{n(S)};$$

that is,

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

Clearly this line of reasoning is valid for the union of any number of nonoverlapping events and thus axiom 3 is also satisfied. ◀

Thus, for any problem in which we are justified in assuming equally likely single-element events, we now have a rule which enables us to compute the probability of occurrence of any event.

Example 2.3.2. Suppose we roll a fair die one time. What is the probability of getting an even number? What is the probability of getting a number which is greater than 4?

Our sample space is $S = \{1, 2, 3, 4, 5, 6\}$. Since the die is fair, we assume the single-element events to be equally likely to occur; each then has probability $\frac{1}{6}$ of occurrence. Let A be the event that an even number occurs and let B be the event that we get a number greater than 4.

$$A = \{2, 4, 6\}, \qquad B = \{5, 6\},$$
$$n(A) = 3, \qquad n(B) = 2, \qquad n(S) = 6,$$

and we get $P(A) = \frac{3}{6}$, $P(B) = \frac{2}{6}$.

Example 2.3.3. We roll a pair of dice one time. What is the probability that the sum of the two numbers is 2? That it is 7? That it is 11?

Our sample space is $S = \{(x_1, x_2): x_1 = 1, 2, \ldots, 6;\ x_2 = 1, 2, \ldots, 6\}$. Since the first die can have a number from 1 through 6 on it and, quite independently, the second die can also have any number 1 through 6 on it, we reason that there are $6 \cdot 6 = 36$ elements belonging to S (or we simply list them all and count them). Thus $n(S) = 36$. Let A be the event that the sum is 2, B the event that the sum is 7, C the event that the sum is 11. Then

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

and we see that $n(A) = 1$, $n(B) = 6$, $n(C) = 2$. Then, since we are assuming that the dice are fair, $P(A) = \frac{1}{36}$, $P(B) = \frac{6}{36}$, $P(C) = \frac{2}{36}$.

Example 2.3.4. A certain class has 20 students. Of the 20, 7 are blue-eyed blond girls, 4 are blue-eyed brunette girls, 5 are blue-eyed blond boys, and the remaining 4 are brown-eyed brunette boys. We select 1 student at random. What is the probability that the selected student is a girl? Has blue eyes? Is a brunette? Is a brown-eyed blond? For ease of notation, we assume that the 20 students are numbered in some specified order. Then our sample space is

$$S = \{1, 2, \ldots, 20\}$$

and the single-element events are equally likely to occur (since the student is selected at random). Define the following events:

A: the selected student is a girl
B: the selected student has blue eyes
C: the selected student is a brunette
D: the selected student is a brown-eyed blond.

Then $N(S) = 20$, $N(A) = 11$, $N(B) = 16$, $N(C) = 8$, $N(D) = 0$, and we have

$$P(A) = \tfrac{11}{20}, \qquad P(B) = \tfrac{16}{20}, \qquad P(C) = \tfrac{8}{20}, \qquad P(D) = 0.$$

Many important problems have equally likely single-element events. As we have seen, computations of probabilities in these cases reduce to counting the number of elements in S and the number of elements in the events of interest. Probabilities are then given by the ratio of these quantities. Since counting the number of elements belonging to an event plays a role in many practical problems, the next two sections are devoted to counting techniques.

EXERCISE 2.3.

1. If two fair coins are flipped, what is the probability that the two faces are alike?
2. If we draw 1 card at random from a standard deck of 52, what is the probability that it is red? That it is a diamond? That it is an ace? That it is the ace of diamonds?
3. Five different colored rubber bowls with the same identical dog food in them are laid out in a row. If a dog chooses a bowl at random from which to eat, what is the probability that he selects the blue one? If a second dog is used, what is the

probability that he chooses the blue one? What is the probability that both choose the blue one?

4. A pair of fair dice is rolled once. Compute the probability that the sum is equal to each of the integers 2 through 12.

5. A one is painted on the head side and a two is painted on the tail side of each of 3 fifty-cent pieces. If the 3 coins are all tossed once (together), compute the probability that the sum of the three numbers occurring is each of the integers 3 through 6.

6. Forty people are riding on the same railroad car. Of this number, 5 are Irish ladies with blue coats, 2 are Irish men with green coats, 1 is an Irish man with a black coat, 7 are Norwegian ladies with brown coats, 2 are Norwegian ladies with blue coats, 6 are Norwegian men with black coats, 4 are German men with green coats, 3 are German ladies with black coats, 5 are German ladies with blue coats, and 5 are German men with black coats. If we select one person at random from this car, what is the probability that the selected person is a man? Is wearing a green coat? Is wearing a brown coat? Is a Norwegian? Is a German? Is a German wearing a green coat?

7. Suppose that a die has been loaded in such a way that the probability of a particular number occurring is proportional to that number. Compute the probabilities of all the single-element events and use these to compute the probability of occurrence of an even number and of a number greater than 4.

2.4. Counting Techniques

As was mentioned in the last section, the solutions of many probability problems depend on being able to count the numbers of elements belonging to particular sets. A number of techniques are invaluable aids in counting the number of elements belonging to certain sets and thus they are of use in solving probability problems in which these sets occur. Unfortunately, there is no general technique that is universally applicable to all counting problems (other than making a complete list and counting the number of items in it); thus it is necessary to attempt to tailor-make counting methods to the particular counting problem in hand, a process that is always frustrating and frequently fruitless. Be that as it may, almost all counting problems seem to yield eventually to a closely reasoned analysis.

The first technique of some generality that we shall discuss is a very simple one frequently referred to as the multiplication principle. This is defined as follows.

DEFINITION 2.4.1. If a first operation can be performed in any of n_1 ways and a second operation can then be performed in any of n_2 ways, both operations can be performed (the second immediately following the first) in $n_1 \cdot n_2$ ways.

This definition can immediately be extended to the simultaneous performance of any number of operations. For example, if we can travel from town A to town B in 3 ways and from town B to town C in 4 ways, then we can travel from A to C in a total of $3 \cdot 4 = 12$ ways. Or, if the operation of tossing a die gives rise to 1 of 6 possible outcomes and the operation of tossing a second die gives rise to 1 of 6 possible outcomes, then the operation of tossing a pair of dice gives rise to $6 \cdot 6 = 36$ possible outcomes.

Example 2.4.1. Suppose that a set A has n_1 elements and a second set B has n_2 elements. Then the Cartesian product $A \times B$ (see Definition 1.3.4) has $n_1 n_2$ elements. The Cartesian product $A \times A$ has n_1^2 elements, $B \times B$ has n_2^2 elements.

DEFINITION 2.4.2. An arrangement of n symbols in a definite order is called a *permutation* of the n symbols.

We shall frequently want to know how many different n-tuples can be made using n different symbols (that is, how many permutations are possible). The multiplication principle will immediately give us the answer. We can count the number of n-tuples by reasoning as follows. In listing all the possible n-tuples, we would perform n natural operations. First we must fill the leftmost position of the n-tuple. Then we must fill the second leftmost position, the third leftmost position, etc. Since we could put any of the n elements available into the leftmost position, this operation can be performed in n ways. After we fill the leftmost position, we can use any of the remaining $n - 1$ elements to fill the second leftmost position; we can use any of the remaining $n - 2$ elements to fill the third leftmost position; and so on until we finally arrive at the rightmost position and want to count the number of ways it can be filled. We have used $n - 1$ elements at this point to fill the first $n - 1$ positions and thus have left only one element which must be used to fill the final position. Then the total number of ways we can perform all n operations (which is also the total number of n-tuples we could make with the given n symbols) is given by the product of the numbers of ways of doing the individual operations. Thus the number of different n-tuples is $n(n-1)(n-2)\cdots 2 \cdot 1$ which we write $n!$ (read n-factorial).

Example 2.4.2. Suppose that the same 5 people park their cars on the same side of the street in the same block every night. How many different orderings of the 5 cars parked on the street are possible? The different orderings of the cars parked on the street could be represented by 5-tuples with 5 distinct elements; then if we can count the number of different 5-tuples that are possible, we also know the number of different orderings on the street. The number of 5-tuples possible is of course $5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$; thus these 5 people could park their cars on the street in a different order every night for 4 months without repeating an ordering they had already used.

DEFINITION 2.4.3. The number of r-tuples we can make ($r \leq n$), using n different symbols (each only once), is called the *number of permutations of n things r at a time* and is denoted by ${}_nP_r$.

How might we compute the value of ${}_nP_r$? Each r-tuple has exactly r positions. The leftmost position could be filled by any of the n symbols; the second leftmost position could then be filled by any of the remaining $n - 1$ symbols, etc. By the time we are ready to fill the rth position, we have used $(r - 1)$ symbols already and any of the remaining $n - (r - 1)$ symbols could be used in the rth position. Thus the total number of r-tuples we could construct is $n(n - 1) \cdots (n - r + 1)$ and we have

$$ {}_nP_r = n(n-1)\cdots(n-r+1). $$

If we multiply this number by $(n - r)!/(n - r)!$, we certainly do not change its value; but in so doing we get a form of ${}_nP_r$ that is much easier to remember:

$$ \begin{aligned} {}_nP_r &= n(n-1)(n-2)\cdots(n-r+1)\frac{(n-r)!}{(n-r)!} \\ &= \frac{n(n-1)(n-2)\cdots(n-r+1)(n-r)(n-r-1)\cdots 2\cdot 1}{(n-r)!} \\ &= \frac{n!}{(n-r)!} \end{aligned} $$

Example 2.4.3. (a) Fifteen cars enter a race. In how many different ways could trophies for first, second, and third place be awarded? This answer is simply ${}_{15}P_3 = 15!/12! = 2730$ since the question is equivalent to asking how many permutations are there of 15 objects, 3 at a time.

(b) How many of the 3-tuples just counted have car number 15 in the first position? This can be answered in two ways. First we might reason that there are ${}_{14}P_2 = 14!/12! = 182$ ways in which the last two positions could be filled, having already put 15 into the first position of the 3-tuple. Alternatively, there must obviously be equal numbers of 3-tuples (in the totality of all possible) having car 15 in first place as there are having car 14 in first place, 13 in first place, etc. Thus, if we divide the total number of 3-tuples by 15, we should get the number that have car 15 in first place; this gives $2730/15 = 182$—the same answer derived above.

Example 2.4.4. (a) How many three-letter "words" can we make using the letters w, i, n, t, e, r (allowing no repetition)? (A "word" is any arrangement of letters, regardless of whether it is in actual fact a word listed in the dictionary for some language.) This is, of course, just ${}_6P_3 = 6!/3! = 120$. The number of four-letter words is ${}_6P_4 = 360$, etc.

(b) Suppose that repetition of a letter is allowed in making three-letter words using w, i, n, t, e, r. How many three-letter words can we make? The answer is $6 \cdot 6 \cdot 6 = 6^3 = 216$, since we now would be able to fill each position with six

letters. The number of four-letter words we could make, allowing repetition, then is $6 \cdot 6 \cdot 6 \cdot 6 = 6^4 = 1296$.

(c) How many three-letter words are there with one or more repeated letters? How many four-letter words are there with one or more repeated letters? We know that there are 216 three-letter words if repetitions are allowed and 120 three-letter words if repetition is not allowed. Thus, there are $216 - 120 = 96$ three-letter words with one or more repeated letters. Analogously, there are $1296 - 360 = 936$ four-letter words with one or more repeated letters.

DEFINITION 2.4.4. The number of subsets, each of size r, that a set with n elements has is called the number of *combinations* of n things r at a time and is denoted by $\binom{n}{r}$.

Remember that sets are not ordered. If we knew $\binom{n}{r}$ and multiplied by $r!$, we would have to get ${}_nP_r$, the number of permutations of n things r at a time, since each distinct subset of r elements would give rise to $r!$ different r-tuples. Thus we get

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

which, dividing through by $r!$, gives us

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

The most difficult part of many counting problems is deciding whether ordering should be of importance. If ordering does not matter we want combinations; if ordering is of importance we want permutations.

Example 2.4.5. (a) How many distinct 5-card hands can be dealt from a standard 52-card deck? Since the 5-card hand remains unchanged if you received the same 5 cards, but in a different order, the answer is

$$ \binom{52}{5} = \frac{52!}{5!\,47!} = 2{,}598{,}960. $$

(b) How many distinct 13-card hands can be dealt from a standard deck?

$$ \binom{52}{13} = 6.35 \times 10^{11}. $$

(c) Suppose that 10 boys go out for basketball at a particular school. How many different teams could be fielded from this school?

$$ \binom{10}{5} = 252. $$

(d) One of the 10 boys out for basketball is named Joe. How many of the 252 teams include Joe as a member? If we want to count only those teams which include Joe, then we need only count how many ways might we select 4 additional individuals to be on the team. This is

$$\binom{9}{4} = 126.$$

$\binom{n}{k}$ is frequently called a combinatorial coefficient. Tables are available giving the value of $\binom{n}{k}$ for varying values of n and k. Let us note a few facts regarding these coefficients. First, $\binom{n}{0} = 1$ for all n as can be seen by simply evaluating the factorials involved. As is also easily seen, $\binom{n}{1} = n$ for all n. One further result which may be useful is the identity $\binom{n}{k} = \binom{n}{n-k}$ for any n and k. By writing these two coefficients out in factorial form, it is immediately evident that they are always equal. Recalling that $\binom{n}{k}$ is the number of subsets of size k that we can construct for a set having n elements, we realize that every time we write down a subset of k items, we leave behind a subset of size $n - k$. Thus it is not surprising that there are always equal numbers of subsets of these two sizes.

Example 2.4.6. (a) How many subsets does a set S with n elements have? (If $\mathcal{F}$ is the class of all subsets of S, then how many elements does $\mathcal{F}$ have?) We are able to compute the answer by using the binomial theorem, reviewed in Appendix 2, and the combinatorial coefficients. Clearly, if we add together the number of subsets of S having $0, 1, 2, \ldots, n$ elements, respectively, this sum would be the total number of subsets of S. $\binom{n}{k}$ is the number of subsets of S, each having k elements, for $k = 0, 1, 2, \ldots, n$. Thus, the total number of subsets is

$$\sum_{k=0}^{n}\binom{n}{k} = \sum_{k=0}^{n}\binom{n}{k}1^k1^{n-k}$$
$$= (1 + 1)^n = 2^n.$$

A set with 2 elements has $2^2 = 4$ subsets; a set with 5 elements has $2^5 = 32$ subsets; a set with 100 elements has $2^{100} = 1.27 \times 10^{30}$ subsets.

(b) If $S = \{1, 2, \ldots, n\}$, how many subsets of S have 1 as an element? Clearly, only one single-element event has 1 as an element. The number of subsets of size 2 having 1 as an element is $\binom{n-1}{1}$ since the second element in the subset could be any one of the remaining $n - 1$. The number of subsets of size 3 having 1 as an element is $\binom{n-1}{2}$; the number of subsets of size r having 1 as an element is $\binom{n-1}{r-1}$ where $r = 1, 2, \ldots, n$. Thus, the total number of subsets having 1 as an element is

$$1 + \binom{n-1}{1} + \binom{n-1}{2} + \cdots + \binom{n-1}{n-1}$$

but, by the binomial theorem, this sum equals 2^{n-1}. Obviously, no matter which single element belonging to S we consider, it belongs to 2^{n-1} events. Since there are 2^n events in total, then exactly half of all possible events occurs, no matter which element belonging to S is the outcome we observe.

EXERCISE 2.4.

1. How many ways can 3 different books be arranged side by side on a shelf?
2. If an item sold by a vending machine costs 40 cents, and the money deposited into the machine must consist of a quarter, a dime, and a nickel, in how many different orders could the money be inserted into the machine?
3. Six people are about to enter a cave in single file. In how many ways could they arrange themselves in a row to go through the entrance?
4. A bag contains 1 red, 1 black, and 1 green marble. I randomly select 1 of the marbles and record its color. I then replace it in the bag, shake the bag, and randomly select a second marble, again recording its color. The second marble is then replaced and a third marble is randomly selected and its color recorded. How many different samples of 3 colors could occur?
5. An ant farm contains both red and black ants. A particular passage in the farm is so narrow that only 1 ant can get through at a time. If 4 ants follow each other through the passage, how many different color patterns (having 4 elements) could be produced (assuming that red ants are indistinguishable from one another, as are black ants)?
6. A particular city is going to give 3 awards to outstanding residents. If 4 people are eligible to receive them, in how many different ways could they be distributed among the 4 people (assuming that no person may receive more than 1 award)?
7. If a set has 3 elements, how many subsets does it have?
8. Could we define a set that has exactly 9 subsets?
9. How many selections of 5 dominoes can be made from a regular 28-domino double-6 set?
10. In how many ways could 2 teams be chosen from an 8-team league?
11. How many committees of 3 people could be chosen from a group of 10?
12. How many 5-man squads could be chosen from a company of 20 men?

13. Given a set of 15 points in a plane, how many lines would be necessary to connect all possible pairs of points?

14. A "complete graph of order 3" is given by connecting 3 points in all possible ways. If 15 points are joined in all possible ways, how many complete graphs of order 3 would be included? Of order $k = 4, 5, 6, \ldots, 15$?

15. Given a box with 2 25-watt, 3 40-watt, and 4 100-watt bulbs, in how many ways could 3 bulbs be selected from the box?

16. Referring to the light bulbs in problem 15, how many of these selections of three bulbs would include both 25-watt bulbs? How many would include no 25-watt bulbs?

17. How many bulb selections defined in problem 15 would include exactly 1 of each of the 3 different wattages?

2.5. Some Particular Probability Problems

In this section we shall take up a few problems that should help acquaint the reader with counting techniques and their applications to probability problems.

Example 2.5.1. A bag contains 4 red and 2 white marbles. If these are randomly laid out in a row, what is the probability that the 2 end marbles are white? That they are not white? That the 2 white marbles are side by side? For convenience we assume that the white marbles are numbered 1 and 2 and that the red marbles are numbered 3 through 6. Then we might adopt as our sample space S the collection of $6! = 720$ permutations of 6 things; i.e.,

$$S = \{(x_1, x_2, \ldots, x_6)\colon x_i = 1, 2, 3, \ldots, 6, \text{ for all } i \text{ and } x_i \neq x_j \text{ for } i \neq j\}.$$

If the marbles are randomly laid out in a row, then each of these 6-tuples is equally likely to occur and we can use our equally likely formula for computing probabilities. Define A to be the event that the first and last marbles are white (the collection of 6-tuples with marbles 1 and 2 on the ends) and B to be the event that marbles 1 and 2 are side by side. Then the number of elements belonging to A is

$$n(A) = 2 \cdot 4! = 48$$

(the white marbles could be on the ends in 2 ways and, for either of these, the red marbles could be arranged between in 4! ways). We also find that

$$n(B) = 5 \cdot 2 \cdot 4! = 240.$$

(There are 5 side-by-side positions for the white marbles to occupy, namely 12, 23, 34, 45, 56; whichever of these is the one to occur, the white marbles can occupy the selected pair of positions in 2 ways and the red balls can be permuted in the remaining positions in 4! ways.) As we noted above,

$$n(S) = 6! = 720$$

and we have

$$P(A) = \tfrac{48}{720} = \tfrac{1}{15}$$

and

$$P(B) = \tfrac{240}{720} = \tfrac{1}{3}.$$

$\bar{A}$ is the event that the white marbles are not on the 2 end positions; we immediately have

$$P(\bar{A}) = 1 - P(A) = \tfrac{14}{15}.$$

An alternative sample space for this problem can be derived as follows. If we pretend that the marbles are going to be put into numbered spots, then all possible outcomes of the experiment can be recorded by keeping track of just the 2 positions that the white marbles occupy; all remaining positions are, of course, filled with the red marbles. All possible pairs of positions are equally likely to occur if the marbles are put down randomly. Thus,

$$S = \{(x_1, x_2)\colon x_1 = 1, 2, \ldots, 6;\ x_2 = 1, 2, \ldots, 6;\ x_1 < x_2\}.$$

Note that S does list every possible pair of position numbers that we could select for the white marbles and that it lists each one only once. The number of elements belonging to S is equal to the number of subsets of size 2 that a set with 6 elements has; i.e.,

$$n(S) = \binom{6}{2} = 15.$$

If we define A and B as above, exactly 1 of these subsets consists of the largest and the smallest elements of S and exactly 5 of them consist of consecutive pairs. Thus,

$$n(A) = 1, \qquad n(B) = 5$$

and, as above,

$$P(A) = \tfrac{1}{15}, \qquad P(\bar{A}) = \tfrac{14}{15}, \qquad P(B) = \tfrac{1}{3}.$$

In many examples, more than one equally likely sample space is possible; used correctly any one of them will give answers to problems of interest.

Example 2.5.2. Suppose that we select a whole number at random between 100 and 999, inclusive. What is the probability that it has at least one 1 in it? What is the probability that it has exactly two 3's in it? For a sample space, we choose

$$S = \{x\colon x = 100, 101, \ldots, 999\}.$$

Then $n(S) = 900$ and, since the number is chosen at random, we assume that all single-element events are equally likely to occur. Define the events:

A: the selected number has at least one 1 in it

B: the selected number has exactly two 3's in it.

We shall find it easy to compute $n(\bar{A})$, then use this to get $P(\bar{A})$, and finally use Theorem 2.2.2 to compute $P(A) = 1 - P(\bar{A})$. We shall compute $n(B)$ directly. The event $\bar{A}$ would be the collection of 3-digit numbers, each of which contains no

1's. The first position of any such number can be filled in 8 ways (since the first digit can be neither 1 nor 0) and each of the succeeding two positions can be filled in 9 ways (since 1 cannot occur in either). Thus,

$$n(\bar{A}) = 8 \cdot 9 \cdot 9 = 648$$

and

$$P(\bar{A}) = \frac{n(\bar{A})}{n(S)} = .72$$

$$P(A) = 1 - P(\bar{A}) = .28.$$

To compute $n(B)$ we can reason as follows: If the first digit is a 3, then one of the succeeding digits must be a 3 and the other can be any of the remaining 9 digits. These two succeeding digits can occur in either of 2 orders, so there are $9 \cdot 2 = 18$ 3-digit numbers having a 3 in the first position and each containing exactly two 3's. If the first digit is not a 3, then this position can be filled in 8 ways (neither 0 nor 3 can be used). The last two positions must then both be filled with 3's. Thus,

$$n(B) = 18 + 8 = 26$$

and

$$P(B) = \frac{n(B)}{n(S)} \doteq .029.$$

Example 2.5.3. Suppose that n people are in a room. If we make a list of their birthdates (month and day of the month), what is the probability that there will be one or more repetitions in the list? (What is the probability that two or more people have the same birthday?) We shall make the assumption that there are only 365 days available for each birthday (ignoring February 29) and that each of these days is equally likely to occur for any individual's birthday. (It can, in fact, be shown that this is the worst possible assumption we might make relative to this probability; that is, if days in March or some other month are more likely to occur as birthdays, then the probability of one or more repeated birthdays is larger than if all days are equally likely.) Our sample space is the collection of all possible n-tuples that could occur for the birthdays, numbering the days of the year sequentially from 1 to 365. Thus,

$$S = \{(x_1, x_2, \ldots, x_n): x_i = 1, 2, \ldots, 365;\ i = 1, 2, \ldots, n\}.$$

The first position in each n-tuple gives the first person's birthday; the second position gives the second person's birthday, etc. Assuming that all days of the year are equally likely for each person's birthday implies that each of these n-tuples is equally likely to occur. By using the counting techniques presented in Section 2.4, we see that

$$n(S) = 365^n.$$

Define A to be the event that there is one or more repetitions of the same number in the n-tuple that occurs. Then $\bar{A}$ is the collection of n-tuples which have no repetitions; we can see rather easily that

$$n(\bar{A}) = 365 \cdot 364 \cdot 363 \cdots (365 - n + 1)$$

which gives us

$$P(\bar{A}) = \frac{n(\bar{A})}{n(S)} = \frac{365 \cdot 364 \cdot 363 \cdots (365 - n + 1)}{365^n}.$$

Again from Theorem 2.2.2,

$$P(A) = 1 - P(\bar{A}).$$

The following table gives the values of $P(\bar{A})$ and $P(A)$ for various values of n. It is somewhat surprising that the probability of a repetition exceeds $\frac{1}{2}$ for as few as 23 people in the room and that for 60 people it is a virtual certainty.

Table 2.1

n	$P(\bar{A})$	$P(A)$
10	.871	.129
20	.589	.411
21	.556	.444
22	.524	.476
23	.493	.507
24	.462	.538
25	.431	.569
30	.294	.706
40	.109	.891
50	.030	.970
60	.006	.994

Example 2.5.4. Suppose that Mrs. Riley claims to be a clairvoyant. Specifically she claims that if she is presented with 8 cards, 4 of which are red and 4 of which are black, she will correctly identify the color of at least 6 of the cards without being able to see their colors. If she is guessing and has no special ability, what is the probability that she would correctly identify at least 6 out of the 8 cards? (She will identify 4 of the cards as red and 4 of the cards as black.) We arbitrarily decide to present her with the 4 red cards first, one by one, and then present her with the 4 black cards. The sample space for the experiment is the set of all possible guesses she might give for the colors of the cards; that is,

$$S = \{(x_1, x_2, \ldots, x_8): x_i = R \text{ or } B, \text{ for all } i;\ \text{exactly } 4x_i\text{'s are } R\}.$$

If she is guessing, then each of the single-element events is equally likely to occur. Since each 8-tuple belonging to S contains exactly 4 R's and exactly 4 B's, we can compute $n(S)$ by counting how many ways we might select 4 positions from the 8 in which to place the R's; thus,

$$n(S) = \binom{8}{4} = 70.$$

Define

A: she identifies at least 6 cards correctly

B: she identifies exactly 6 cards correctly

C: she identifies all 8 cards correctly.

Since she will call 4 cards red and 4 cards black, it is not possible for her to be correct on exactly 7 cards; thus,

$$A = B \cup C$$

and, since

$$B \cap C = \varnothing,$$

$$P(A) = P(B) + P(C).$$

Clearly, $n(C) = 1$ so $P(C) = \frac{1}{70}$. If B is to occur, she must identify exactly 3 of the 4 red cards correctly and exactly 3 of the 4 black cards correctly. The 1 red card that she is wrong on could be any of the 4 and the 1 black card she is wrong on could be any one of the 4. Thus, the number of 8-tuples having 1 B in the first 4 positions and 1 R in the last 4 positions is

$$n(B) = 4 \cdot 4 = 16$$

and we have

$$P(B) = \tfrac{16}{70}.$$

Thus

$$P(A) = \tfrac{1}{70} + \tfrac{16}{70} = \tfrac{17}{70} \doteq .243;$$

if she only guesses, there is slightly less than 1 chance in 4 of her doing as well as she claims.

EXERCISE 2.5.

1. In how many ways could a dozen oranges be chosen from a table holding 30 oranges? (How many distinct collections of 12 oranges could be made?)

2. How many arrangements could be made of 5 red balls and 1 orange ball?

3. In how many of the arrangements counted in problem 2 are the red balls all together?

4. A certain market uses red boxes and green boxes for displays at Christmas time. In how many ways could the market arrange 20 boxes in a row if 15 of them are red and 5 are green? If there are 10 boxes of each color?

5. Ten people in total are nominated for a slate of 3 offices. If every group of 3 people has the same probability of winning, what is the probability that a particular person will be on the winning slate? That a particular pair of people will be on the winning slate?

6. Two people are to be selected at random to be set free from a prison with a population of 100. What is the probability that the oldest prisoner is 1 of the 2 selected? That the oldest and youngest are the pair selected?

7. In a certain national election year, governors were to be elected in 30 states. Assume that in every state there were only 2 candidates (called the Republican and Democratic candidates, respectively). What is the probability that the Republicans carried all 30 states, assuming that each state was equally likely to elect either party? What is the probability that the same party carried all the states?

8. A, B, and C are going to race. What is the probability that A will finish ahead of C, given that all are of equal ability (and no ties can occur)? What is the probability that A will finish ahead of both B and C?

9. Each of 5 people is asked to distinguish between vanilla ice cream and French vanilla custard (each is given a small sample of both and asked to identify which is ice cream). If all 5 people are guessing, what is the probability that all will correctly identify the ice cream? If all 5 are guessing, what is the probability that at least 4 will identify the ice cream correctly?

10. Compute the probability that a group of 5 cards drawn at random from a 52-card deck will contain

(a) exactly 2 pair
(b) a full house (3 of one denomination and 2 of another)
(c) a flush (all 5 from the same suit)
(d) a straight (5 in sequence, beginning with ace or deuce or trey, . . . , or ten).

11. n people are in a room. Compute the probability that at least 2 have the same birth month. Evaluate this probability for $n = 3, 4, 5, 6$.

12. A student is given a true-false examination with 10 questions. If he gets 8 or more correct, he passes. If he is guessing, what is his probability of passing the examination?

13. Eight black and 2 red balls are randomly laid out in a row. What is the probability that the 2 red balls are side by side? That the 2 red balls are occupying the end positions?

14. A person is to be presented with 3 red and 3 white cards in a random sequence. He knows that there will be 3 of each color; thus he will identify 3 cards as being of each color. If he is guessing, what is his probability of correctly identifying all 6 cards? Of identifying exactly 5 correctly? Exactly 4?

2.6. Conditional Probability

In some applications we shall be given the information that an event A occurred and will be asked the probability that another event B also occurred. For example, we might be given that the card we selected from a regular 52-card bridge deck was red and then might want to know the probability that the card selected was the ace of hearts. Or, when running an opinion poll, we might be given the fact that the person we have selected is a Republican and then might also want to know the probability that he favors our current actions in Viet Nam. Or, when conducting a medical research experiment, we might be given that a randomly selected person has a family history of diabetes and then ask what is the probability that this person also has diabetes. In each of these examples we are given that an event A has occurred and want to know the probability that a second event B also has occurred.

If we are given that A has already occurred, then in effect we now have the event A as our sample space since we know that any element $x \in \bar{A}$ did not occur. Thus it would seem reasonable to measure the probability that B also has occurred by the relative proportion of the time that A and B occur together (relative to the total probability of A occurring). This is in fact the definition of the conditional probability of B occurring, given that A has occurred, as we see in the following definition.

DEFINITION 2.6.1. The *conditional probability* of B occurring, given that A has occurred (written $P(B \mid A)$) is $P(B \mid A) = P(B \cap A)/P(A)$ if $P(A) > 0$. If $P(A) = 0$, we define $P(B \mid A) = 0$.

Example 2.6.1. We roll a pair of fair dice 1 time and are given that the 2 numbers that occur are not the same. Compute the probability that the sum is 7 or that the sum is 4 or that the sum is 12. Define the event

A: the two numbers that occur are different.

Then we are given that A has occurred. Also define the events

B: the sum is 7

C: the sum is 4

D: the sum is 12.

Then, assuming equally likely single-element events, we find that

$$P(A) = \tfrac{5}{6}, \quad P(B) = \tfrac{1}{6}, \quad P(C) = \tfrac{1}{12}, \quad P(D) = \tfrac{1}{36},$$
$$P(A \cap B) = \tfrac{1}{6}, \quad P(A \cap C) = \tfrac{1}{18}, \quad P(A \cap D) = 0$$

and thus we have

$$P(B \mid A) = \frac{\frac{1}{6}}{\frac{5}{6}} = \tfrac{1}{5}$$

$$P(C \mid A) = \frac{\frac{1}{18}}{\frac{5}{6}} = \tfrac{1}{15}$$

$$P(D \mid A) = \frac{0}{\frac{5}{6}} = 0.$$

Example 2.6.2. From past experience with the illnesses of his patients, a doctor has gathered the following information: 5% feel that they have cancer and do have cancer, 45% feel that they have cancer and don't have cancer, 10% do not feel that they have cancer and do have it, and finally 40% feel that they do not have cancer and really do not have it. These percentage figures then imply the following probabilities for a randomly selected patient from this doctor's practice. Define the events

A: the patient feels he has cancer

B: the patient has cancer.

Then

$$P(A \cap B) = .05, \qquad P(A \cap \bar{B}) = .45,$$
$$P(\bar{A} \cap B) = .1, \qquad P(\bar{A} \cap \bar{B}) = .4.$$

Then we have

$$P(A) = P(A \cap B) + P(A \cap \bar{B}) = .5$$
$$P(B) = P(A \cap B) + P(\bar{A} \cap B) = .15.$$

The probability that a patient has cancer then, given that he feels he has it, is

$$P(B \mid A) = \frac{.05}{.5} = .1.$$

The probability he has cancer given that he does not feel that he has it is

$$P(B \mid \bar{A}) = \frac{.1}{.5} = .2.$$

The probability he feels he has cancer given that he does not have it is

$$P(A \mid \bar{B}) = \frac{.45}{.85} = \frac{9}{17}.$$

The probability he feels he has cancer given that he does have it is

$$P(A \mid B) = \frac{.05}{.15} = \frac{1}{3}.$$

One of the major uses of conditional probability is to allow easy computation of the probabilities of intersections of events for certain experiments. The equation

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

implies that

$$P(B \cap A) = P(A \cap B) = P(A)P(B \mid A). \tag{2.1}$$

The equation

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

implies that

$$P(A \cap B) = P(B)P(A \mid B). \tag{2.2}$$

One or the other of these formulas for $P(A \cap B)$ is useful in many problems.

Example 2.6.3. We select 2 balls at random without replacement from an urn which contains 4 white and 8 black balls. (a) Compute the probability that both balls are white. (b) Compute the probability that the second ball is white.

(a) Define

A: the first ball is white

B: the second ball is white

C: both balls are white.

Then

$$A \cap B = C \text{ and } P(C) = P(A \cap B) = P(A)P(B \mid A)$$
$$= \tfrac{4}{12} \cdot \tfrac{3}{11} = \tfrac{1}{11}.$$

(b) Clearly $B = (A \cap B) \cup (\bar{A} \cap B)$ and

$$(A \cap B) \cap (\bar{A} \cap B) = \varnothing.$$

Then

$$\begin{aligned} P(B) &= P(A \cap B) + P(\bar{A} \cap B) \\ &= P(A)P(B \mid A) + P(\bar{A})P(B \mid \bar{A}) \\ &= \tfrac{4}{12} \cdot \tfrac{3}{11} + \tfrac{8}{12} \cdot \tfrac{4}{11} \\ &= \tfrac{1}{3}. \end{aligned}$$

Notice that the probability of drawing a white ball the second draw (even though the first one is not replaced) is $\frac{1}{3}$, the same as the probability that the first ball drawn is white. This can be shown to be the case generally. $P(B)$ is an unconditional probability and, as indicated by the way it was computed above, it is an average of the two conditional probabilities $P(B \mid A)$ and $P(B \mid \bar{A})$.

Example 2.6.4. Box 1 contains 4 defective and 16 nondefective light bulbs. Box 2 contains 1 defective and 1 nondefective light bulb. We roll a fair die 1 time. If we get a 1 or a 2, then we select a bulb at random from box 1. Otherwise we select a bulb from box 2. What is the probability that the selected bulb will be defective? Define

A: we select a bulb from box 1

B: the selected bulb is defective.

Then $P(A) = \frac{1}{3}$, $P(\bar{A}) = \frac{2}{3}$, $P(B \mid A) = \frac{1}{5}$, and $P(B \mid \bar{A}) = \frac{1}{2}$. Since

$$B = (B \cap A) \cup (B \cap \bar{A})$$

and

$$(B \cap A) \cap (B \cap \bar{A}) = \varnothing,$$

we have

$$\begin{aligned} P(B) &= P(B \cap A) + P(B \cap \bar{A}) \\ &= P(A)P(B \mid A) + P(\bar{A})P(B \mid \bar{A}) \\ &= \tfrac{1}{3} \cdot \tfrac{1}{5} + \tfrac{2}{3} \cdot \tfrac{1}{2} = \tfrac{6}{15} = \tfrac{2}{5}. \end{aligned}$$

The following result, frequently called Bayes' theorem or Bayes' formula, is useful in many applied problems. It is named after the Reverend Thomas

Bayes, one of the early writers on probability theory; it has recently been applied in many different sorts of problems and plays an important role in many branches of applied statistics.

Theorem 2.6.1. (Bayes) Suppose that we are given k events $A_1, A_2, \ldots, A_k$ such that:

1. $A_1 \cup A_2 \cup \cdots \cup A_k = S$
2. $A_i \cap A_j = \varnothing, \qquad \text{for all } i \neq j$

(these events form a *partition* of S); then for any event $E \subset S$,

$$P(A_j \mid E) = \frac{P(A_j)P(E \mid A_j)}{\sum_{i=1}^{k} P(A_i)P(E \mid A_i)}, \qquad j = 1, 2, \ldots, k.$$

Proof: Because of properties 1 and 2 listed above, we know that for any event $E \subset S$ we can write

$$E = (E \cap A_1) \cup (E \cap A_2) \cup \cdots \cup (E \cap A_k)$$

where

$$(E \cap A_i) \cap (E \cap A_j) = \varnothing, \qquad \text{for all } i \neq j.$$

Thus

$$\begin{aligned} P(E) &= P(E \cap A_1) + P(E \cap A_2) + \cdots + P(E \cap A_k) \\ &= P(A_1)P(E \mid A_1) + P(A_2)P(E \mid A_2) + \cdots + P(A_k)P(E \mid A_k) \end{aligned}$$

by Equation 2.1. By definition

$$P(A_j \mid E) = \frac{P(A_j \cap E)}{P(E)}, \qquad j = 1, 2, \ldots, k.$$

And by using Equation 2.1 and the result above for $P(E)$, we have

$$P(A_j \mid E) = \frac{P(A_j)P(E \mid A_j)}{\sum_{i=1}^{k} P(A_i)P(E \mid A_i)}$$

which is the desired result. ◀

Example 2.6.5. Suppose that you are a political prisoner in Russia and are to be exiled in one of two places: Siberia or Mongolia. The probabilities of being sent to these two places are .7 and .3, respectively. It is also known that if you randomly select a resident of Siberia, the probability that he will be wearing a seal-skin coat is .8, whereas this same event has probability .4 in Mongolia. Late one night you are blindfolded and thrown on a truck. Two weeks later (you estimate) the truck stops, you are told you have arrived at your place of exile, and your blindfold is removed. The first person you see is not wearing a seal-skin coat. What is the probability that your place of exile is Siberia? Bayes theorem can be used to answer this question. Define

$$A\text{: you are sent to Siberia.}$$

Then we have

$$\bar{A}\text{: you are sent to Mongolia.}$$

Also define

$$B\text{: randomly selected resident wears seal-skin coat.}$$

The information we have been given then is:

$$P(A) = .7, \quad P(\bar{A}) = .3, \quad P(B \mid A) = .8, \quad P(B \mid \bar{A}) = .4$$

and we are asked to compute $P(A \mid \bar{B})$. Since A and $\bar{A}$ form a partition of the sample space we have, from Bayes theorem,

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

$$= \frac{(.7)(.2)}{(.7)(.2) + (.3)(.6)} = \frac{7}{16}.$$

EXERCISE 2.6.

1. An urn contains 4 balls numbered 1, 2, 3, 4, respectively. Two balls are drawn without replacement. Let A be the event that the sum is 5 and let B_i be the event that the first ball drawn has an i on it, $i = 1, 2, 3, 4$. Compute $P(A \mid B_i)$, $i = 1, 2, 3, 4$, and $P(B_i \mid A)$, $i = 1, 2, 3, 4$.

2. Suppose that the two balls of problem 1 are drawn with replacement. Let A and B_i be defined as above and compute $P(A \mid B_i)$ and $P(B_i \mid A)$, $i = 1, 2, 3, 4$.

3. A fair coin is flipped 4 times. What is the probability that the fourth flip is a head, given that each of the first 3 flips resulted in heads?

4. A fair coin is flipped 4 times. What is the probability that the fourth flip is a head, given that 3 heads occurred in the 4 flips? Given that 2 heads occurred in the 4 flips?

5. Urn 1 contains 2 red and 4 blue balls, urn 2 contains 10 red and 2 blue balls. If an urn is chosen at random and a ball is removed from the chosen urn, what is the probability that the selected ball is blue? That it is red?

6. Suppose that in problem 7, instead of selecting an urn at random we roll a die and select the ball from urn 1 if a 1 occurs on the die and otherwise select the ball from urn 2. What is the probability that the selected ball is blue? That it is red?

7. Five cards are selected at random without replacement from a 52-card deck. What is the probability that they are all red? That they are all diamonds?

8. An urn contains 2 red, 2 white, and 2 blue balls. Two balls are selected at random without replacement from the urn. Compute the probability that the second ball drawn is red.

9. An urn contains 2 black and 5 brown balls. A ball is selected at random. If the ball drawn is brown, it is replaced and two additional brown balls are also put into the urn. If the ball drawn is black, it is not replaced in the urn and no additional balls are added. A ball is then drawn from the urn the second time. What is the probability that it is brown?

10. The two-stage experiment described in problem 9 was performed and we are given that the ball selected at the second stage was brown. What is the probability that the ball selected at the first stage was also brown?

11*. Suppose that medical science has a cancer-diagnostic test that is 95% accurate both on those that do have cancer and on those that do not have cancer. If .005 of the population actually does have cancer, compute the probability that a particular individual has cancer, given that the test says he has cancer.

12. In a large midwestern school, 1% of the student body participates in the intercollegiate athletic program; 10% of these people have a grade point of 3 or more (out of 4) whereas 20% of the remainder of the student body have a grade point of 3 or more. What proportion of the total student body have a grade point of 3 or more? Suppose we select 1 student at random from this student body and find that he has a grade point of 3.12. What is the probability that he participates in the intercollegiate athletic program?

2.7. Independent Events

It is possible to define and be interested in events A and B such that if we know that A has occurred, then we are also certain that B has occurred (take any example where $A \subset B$). In such a case there is certainly a degree of dependence between A and B. As we shall see in this section, it is also possible to have two events A and B such that the knowledge that A has occurred gives no information on whether or not B also has occurred. Two such events will be called independent (also referred to as being statistically independent). The formal definition of independent events is as follows.

DEFINITION 2.7.1. Two events, A and B, are *independent* if and only if

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

Before looking at some examples of independent events, let us note that if A and B are independent, then

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

Thus, the conditional probability of A occurring is the same as the unconditional probability of A. The knowledge that B occurred didn't change the

* Adapted from Emanuel Parzen, *Modern Probability Theory and Its Applications*, Wiley, 1960.

probability of A's occurrence. Similarly, it is easily seen that $P(B \mid A) = P(B)$ if A and B are independent.

Example 2.7.1. Assume that the numbers given in the cells of Table 2.2 give the probabilities of a randomly selected individual falling into the given cell. That is,

Table 2.2

	Gets Cancer	Does Not Get Cancer
Smoker	.50	.20
Nonsmoker	.10	.20

if we let A be the event that the selected individual is a smoker and let B be the event that the selected individual gets cancer, then

$$P(A \cap B) = .5, \quad P(A \cap \bar{B}) = .2, \quad P(\bar{A} \cap B) = .1$$

and

$$P(\bar{A} \cap \bar{B}) = .2.$$

Since

$$P(A) = P(A \cap B) + P(A \cap \bar{B}) = .7$$
$$P(B) = P(A \cap B) + P(\bar{A} \cap B) = .6,$$

We see that $P(A \cap B) = 2 \neq (.7)(.6)$, so A and B are not independent.

Example 2.7.2. If 2 fair dice are rolled 1 time, show that the 2 events

A: the sum of the 2 dice is 7

B: the 2 dice have the same number

are not independent. As we have seen before, $P(A) = P(B) = \frac{1}{6}$; $A \cap B = \varnothing$ so $P(A \cap B) = 0$ which is not $\frac{1}{6} \cdot \frac{1}{6}$. Thus, the 2 events are not independent.

Two events which cannot happen simultaneously are said to be mutually exclusive, as is given in the next definition.

DEFINITION 2.7.2. A and B are *mutually exclusive* if and only if $A \cap B = \varnothing$.

The definitions of independence and mutually exclusive are frequently confused; this is probably caused by the fact that in common English usage the word independent is frequently used to mean "having nothing to do with." The phrase "having nothing to do with" could be interpreted to mean that two events could not happen together, which we have just defined as being mutually exclusive. The two definitions (2.7.1 and 2.7.2) are certainly not the same in content, as we can see from the following theorem.

Theorem 2.7.1. Assume that $P(A) \neq 0$ and $P(B) \neq 0$. Then A and B independent implies that they are not mutually exclusive and A and B mutually exclusive implies that they are not independent.

Proof: Suppose that A and B are independent. Then $P(A \cap B) = P(A)P(B) \neq 0$, since $P(A) \neq 0$ and $P(B) \neq 0$. Thus they are not mutually exclusive. Now suppose A and B are mutually exclusive. Then $A \cap B = \varnothing$ and $P(A \cap B) = 0$. But since $P(A) \neq 0$ and $P(B) \neq 0$, $P(A)P(B) \neq 0$ and they are then not independent. ◀

Independence of three events is defined as follows.

DEFINITION 2.7.3. A, B, and C are independent if and only if:

1. $P(A \cap B) = P(A)P(B)$
2. $P(A \cap C) = P(A)P(C)$
3. $P(B \cap C) = P(B)P(C)$
4. $P(A \cap B \cap C) = P(A)P(B)P(C)$.

Many examples can be given which show that the first three of these conditions do not imply the fourth and vice versa. The following is an example in which the first three equations hold but the fourth does not.

Example 2.7.3. Suppose that we have a bowl with 4 tags in it labelled 000, 110, 101, 011, respectively. We draw 1 tag from the bowl at random and define A_i ($i = 1, 2, 3$) to be the event that a 0 occurs in the ith position. Then equations 1, 2, and 3 in Definition 2.7.3 are satisfied by these three events, but equation 4 is not. A_1, A_2, A_3 are called pairwise independent events but not independent events. Problem 6 below shows a case in which equation 4 is satisfied but at least one of 1, 2, and 3 is not.

Many practical experiments consist of independent trials; by independent trials we mean that a particular outcome in one trial has no affect on the outcome observed in another trial. This idea is made more exact in the following definition.

DEFINITION 2.7.4. An experiment consists of n *independent trials* if and only if: (1) S is the Cartesian product of n sets $S_1, S_2, \ldots, S_n$, and (2) the probability of occurrence of a single-element event $A \subset S$ is the product of the probabilities of occurrence of appropriate single-element events $A_i \subset S_i$, $i = 1, 2, \ldots, n$; that is,

$$P(A) = P_1(A_1)P_2(A_2) \cdots P_n(A_n)$$

where $A \subset S$, $A_i \subset S_i$, $i = 1, 2, \ldots, n$ and $A, A_1, \ldots, A_n$ are each single-element events.

Note immediately then that an experiment that consists of n independent trials has n-tuples as elements of its sample space. Furthermore, probabilities of single-element events are assigned in a special way; this special way in fact gives an easy method of computing many probabilities.

Example 2.7.4. A box contains 15 electron tubes of which 4 are defective. We select 5 tubes at random, with replacement, from the box. We choose as our sample space

$$S = \{(x_1, x_2, \ldots, x_5): x_i = n \text{ or } d, \quad i = 1, 2, \ldots, 5\},$$

where n is meant to stand for nondefective and d for defective. A single-element event then is a 5-tuple; in fact, if we define

$$S_i = \{n, d\}, \qquad i = 1, 2, \ldots, 5,$$

we see that

$$S = S_1 \times S_2 \times S_3 \times S_4 \times S_5.$$

Since the sampling is done with replacement, the composition of the box is the same for each of the draws. Thus, if the first tube drawn is defective, this has no effect on whether the second tube is also defective (or the third, etc.). The draws then are independent and this is an experiment with 5 independent trials. For each S_i we see that

$$P_i(\{n\}) = \tfrac{11}{15}, \qquad P_i(\{d\}) = \tfrac{4}{15}.$$

This then gives us the assignment of probabilities for each single-element event of S. For example

$$P(\{(n, n, n, n, n)\}) = \tfrac{11}{15} \cdot \tfrac{11}{15} \cdot \tfrac{11}{15} \cdot \tfrac{11}{15} \cdot \tfrac{11}{15} = (\tfrac{11}{15})^5$$
$$P(\{(d, d, d, d, d)\}) = \tfrac{4}{15} \cdot \tfrac{4}{15} \cdot \tfrac{4}{15} \cdot \tfrac{4}{15} \cdot \tfrac{4}{15} = (\tfrac{4}{15})^5$$
$$P(\{(n, d, n, d, n)\}) = \tfrac{11}{15} \cdot \tfrac{4}{15} \cdot \tfrac{11}{15} \cdot \tfrac{4}{15} \cdot \tfrac{11}{15} = (\tfrac{4}{15})^2(\tfrac{11}{15})^3$$
$$P(\{(d, d, n, n, n)\}) = \tfrac{4}{15} \cdot \tfrac{4}{15} \cdot \tfrac{11}{15} \cdot \tfrac{11}{15} \cdot \tfrac{11}{15} = (\tfrac{11}{15})^3(\tfrac{4}{15})^2$$

and so on. Once we know the probabilities of the single-element events, we can use these to compute the probabilities of any other events of interest. Note that if we sampled tubes from this box without replacement, we would not have independent draws because the number of tubes to be drawn from would not be the same for each draw nor would the relative proportion of defective tubes remain constant.

Example 2.7.5. The Iowa State University football team plays 11 different teams on successive Saturdays. If we assume that their performances from one Saturday to the next are independent, then the experiment which consists of their full schedule for a year is made up of 11 independent trials. Define

$$S_i = \{W, L, T\}, \qquad i = 1, 2, \ldots, 11,$$

where W stands for win, L for loss, and T for tie, and the sample space S for the year's games can be written

$$S = S_1 \times S_2 \times \cdots \times S_{11}.$$

Then the probability of occurrence of any single-element event belonging to S is defined to be the product of the probabilities of the single-element events for the appropriate trials; that is

$$P(\{(x_1, x_2, \ldots, x_{11})\}) = P_1(\{x_1\})P_2(\{x_2\}) \cdots P_{11}(\{x_{11}\})$$

where

$$(x_1, x_2, \ldots, x_{11}) \in S, \qquad x_1 \in S_1, x_2 \in S_2, \ldots, x_{11} \in S_{11}.$$

The differing subscript on P for the various trials is meant to indicate that the probability of Iowa State winning (or losing or tying) a game is not necessarily the same for all weeks. The probability that they win all their games is then

$$\prod_{i=1}^{11} P_i(\{W\});$$

the probability that they lose all their games is

$$\prod_{i=1}^{11} P_i(\{L\}).$$

(The product notation is discussed in Appendix 1.)

EXERCISE 2.7.

1. One fair coin is flipped 2 times. Are the 2 events

   A: a head occurs on the first flip

   B: a head occurs on the second flip

   independent?

2. A fair coin is flipped 2 times. Let A be the event that a head occurs on the first flip and let B be the event that the same face does not occur on both flips. Are A and B independent?

3. An urn contains 4 balls numbered 1, 2, 3, 4, respectively. Two balls are drawn without replacement. Let A be the event that the first ball drawn has a 1 on it and let B be the event that the second ball has a 1 on it. Are A and B independent?

4. If the drawing is done with replacement in problem 3, are A and B independent?

5. A pair of dice is rolled 1 time. Let A be the event that the first die has a 1 on it, B the event that the second die has a 6 on it, and C the event that the sum is 7. Are A, B, and C independent?

6. A fair coin is flipped 3 times. Let A be the event that a head occurs on the first flip, let B be the event that at least 2 tails occur, and let C be the event that we get exactly 1 head or that we get tail, head, head in that order. Show that these 3 events satisfy equation 4 of Definition 2.7.3. but not equations 1, 2, or 3.

7. Prove that if A and B are independent, so are $\bar{A}$ and $\bar{B}$.

8. The probability that a certain basketball player scores on a free throw is .7. If in a game he gets 15 free throws, compute the probability that he makes them all. Compute the probability that he makes 14 of them. What assumptions have you made in deriving your answer?

9. Three teams, A, B, and C, enter a round-robin tournament. (Each team plays 2 games, 1 against each of the possible opponents. The winner of the tournament, if

there is a winner, is the team winning both its games.) Assume that the game played is one in which a tie is not allowed. We assume the following probabilities:

$$P(A \text{ beats } B) = .7$$
$$P(B \text{ beats } C) = .8$$
$$P(C \text{ beats } A) = .9.$$

Compute the probability that team A wins the tournament; that team B wins the tournament. Compute the probability no one wins the tournament.

2.8. Discrete and Continuous Sample Spaces

The reader may have noted that the sample spaces used in all the preceding examples were finite sets. Such sample spaces are special cases of what are called discrete sample spaces. Discrete sample spaces are defined as follows.

DEFINITION 2.8.1. A *discrete* sample space is one which has a finite or a countably infinite number of elements.

Generally, discrete sample spaces are those for which it is meaningful to consider single-element events. That is, knowledge of the probabilities of occurrence of the single-element events is sufficient for computing the probabilities of any event. All of the finite sample spaces we have seen thus far have been of this sort. The following example gives a problem in which the sample space is countably infinite, yet we still are able to compute probabilities from knowledge of the single-element events.

Before looking at this example, let us note what our third axiom is saying when we have an infinite sample space. In an infinite sample space it is possible to have an infinite number of nonoverlapping events. (Consider, for example, the single-element events.) Axiom 3 says then that the probability assigned to the union of an infinite number of nonoverlapping events must be equal to the value to which an infinite series converges (the sum of the individual probabilities assigned to the events). Appendix 3 gives a more complete discussion of orders of infinity.

Example 2.8.1. Suppose we flip a fair coin until we get a head. Compute the probability that it takes less than 4 flips to conclude the experiment. Compute the probability that it takes an even number of flips to conclude the experiment. Our sample space is

$$S = \{H, TH, TTH, \ldots\}.$$

H signifies that we get a head on the first flip, TH that we first got a head on the second flip, TTH that we first got a head on the third flip, etc. Clearly we can set up a one-to-one correspondence between the elements of S and the positive integers; thus S has a countably infinite number of elements and is a discrete sample space. Assuming that the coin is fair and that the flips are independent, we can easily assign

probabilities to the single-element events as follows.

$$P(\{H\}) = \tfrac{1}{2}, \qquad P(\{TH\}) = \tfrac{1}{2}\cdot\tfrac{1}{2} = \tfrac{1}{4},$$
$$P(\{TTH\}) = \tfrac{1}{2}\cdot\tfrac{1}{2}\cdot\tfrac{1}{2} = \tfrac{1}{8}, \qquad P(\{TTTH\}) = \tfrac{1}{2}\cdot\tfrac{1}{2}\cdot\tfrac{1}{2}\cdot\tfrac{1}{2} = \tfrac{1}{16},$$

etc. First of all we should make sure that this assignment satisfies axiom 1. Clearly

$$\begin{aligned} P(S) &= \tfrac{1}{2} + \tfrac{1}{4} + \tfrac{1}{8} + \tfrac{1}{16} + \cdots \\ &= \tfrac{1}{2}(1 + \tfrac{1}{2} + \tfrac{1}{4} + \tfrac{1}{8} + \cdots) \\ &= \tfrac{1}{2}\cdot\frac{1}{1-\frac{1}{2}} = 1, \end{aligned}$$

since the parenthetic expression is just a geometric progression with $r = \tfrac{1}{2}$. Thus we do satisfy axiom 1. Define

A: it takes less than 4 flips to conclude the experiment

B: it takes an even number of flips to conclude the experiment.

That is,

$$A = \{H, TH, TTH\}$$
$$B = \{TH, TTTH, TTTTTH, \ldots\}.$$

Then

$$\begin{aligned} P(A) &= \tfrac{1}{2} + \tfrac{1}{4} + \tfrac{1}{8} = \tfrac{7}{8} \\ P(B) &= \tfrac{1}{4} + \tfrac{1}{16} + \tfrac{1}{64} + \tfrac{1}{256} + \cdots \\ &= \tfrac{1}{4}(1 + \tfrac{1}{4} + \tfrac{1}{16} + \tfrac{1}{64} + \cdots) \\ &= \tfrac{1}{4}\cdot\frac{1}{1-\frac{1}{4}} = \tfrac{1}{3}. \end{aligned}$$

Example 2.8.2. Suppose that we select a number at random from the positive integers. What is the probability that it is even? The sample space for this experiment is

$$S = \{1, 2, 3, \ldots\},$$

so S is discrete. However, it is not possible to assign the same probability to each of the single-element events. No matter how small the value of this common probability, the sum of the probabilities of the single-element events is infinite. That is, suppose we say

$$P(\{x\}) = p > 0 \qquad \text{for } x = 1, 2, 3, \ldots$$

Then

$$P(S) = \sum_{x=1}^{\infty} P(\{x\}) = \sum_{x=1}^{\infty} p$$

and this sum diverges, no matter how close to 0 p is. Thus it would appear that there is no way we could satisfactorily describe such an experiment. This conclusion is perfectly correct because, after a little reflection, we would have to conclude that there is absolutely no way in which we could perform the experiment. No matter what physical mechanism we used to "select a number at random from the positive integers," there are too many positive integers to be considered since the integers have no end. The experiment itself makes no sense; this is why we cannot describe it. In spite of the foregoing, it would seem reasonable to give $\frac{1}{2}$ as the probability of selecting an even number since every other integer is even. This conclusion is actually based on reasoning such as the following. If we were to select randomly an integer from the set $\{1, 2, 3, \ldots, M\}$ where M is very large, the probability that the number is even is $\frac{1}{2}$ if M is even and $\frac{1}{2} - \dfrac{1}{2M}$ if M is odd. In either case, it is about $\frac{1}{2}$ and, as M gets bigger and bigger, it gets closer and closer to $\frac{1}{2}$ (for odd M; for even M it is $\frac{1}{2}$).

Many examples give rise to continuous sample spaces. These are defined as follows:

DEFINITION 2.8.2. A *continuous* (one-dimensional) sample space is one which has as elements all of the points in some interval on the real line.

Thus the sets

$$\{x: 0 < x < 1\}, \qquad \{x: 10 \leq x \leq 20\}, \qquad \{x: x > 0\}$$

are all examples of continuous sample spaces. Generally, if the experiment consists of observing something which could lie anywhere along a continuous line, we shall want to use a continuous sample space. Again, subsets of the continuous interval are events; the single-element events would be the sets of single points in the interval. In the case of continuous sample spaces, however, we shall generally have to conclude that the probabilities of occurrence of these individual points must be zero. Otherwise, since there are a noncountable infinity of them in any interval, we would not be able to satisfy our first axiom. Thus, in continuous sample spaces, we cannot compute the probabilities of occurrence of any event from knowledge of the probabilities of the single-element events.

Another difficulty is also encountered. In discrete sample spaces, all subsets are called events and probabilities can consistently be assigned to them. It can be shown that in a continuous sample space S, not all subsets are probabilizable; that is, certain examples of subsets can be constructed such that any assignment of probabilities to them is inconsistent with our three axioms. Such subsets are not called events and thus are not assigned probabilities. The reader should rest assured, however, that in any practical problem we shall be able to call any outcomes of interest events. In more advanced courses it is shown that the subsets of S belonging to the class of Borel sets are all probabilizable; furthermore, the class of Borel sets includes all of the continuous-interval subsets of S, as well as unions and intersections of such continuous-interval subsets. Without going further into the matter,

we merely acknowledge the fact that not all subsets of S are called events if S consists of all the points in some continuous interval; those that are events are actually Borel sets.

Continuous sample spaces, in quite general examples, are most easily studied after we are acquainted with random variables (to be introduced in Chapter 3). Let us content ourselves here with looking at a particular type of continuous sample space and the way in which we assign probabilities to events. The particular type of continuous sample space we shall investigate now is one which has equally likely elements. As we have seen, each of the single-element events (single points) must then have probability 0. Therefore, what we really mean by saying that the elements of S are equally likely to occur is that the probability of the occurring point lying in a continuous subinterval of S is proportional to the length of the subinterval. If, for example, A and B are both continuous subintervals of S and if A is twice as long as B, then the probability of A containing the outcome is twice as big as the probability of B containing the outcome.

It can be shown that for continuous sample spaces, we must specify a rule for assigning probabilities to continuous subintervals rather than specifying probabilities for all single-element events. In this section, as has been noted, we shall be concerned only with the particular continuous sample space which has equally likely outcomes. As mentioned above, this implies that the probability of the outcome lying in a continuous subinterval should be proportional to the length of the subinterval. Accordingly, suppose we are given an experiment whose outcome is equally likely to lie anywhere in the real interval from a to b inclusive. Then

$$S = \{x: a \leq x \leq b\}.$$

Suppose A is a continuous subinterval of S; i.e.,

$$A = \{x: c \leq x \leq d\},$$

where $c > a$, $d < b$. (See Figure 2.3.) For any continuous subinterval A, define the set function

$$L(A) = \text{length of } A.$$

Then, for the particular subset defined above we have

$$L(A) = d - c,$$

and for S we have

$$L(S) = b - a.$$

One rule for assigning probabilities then is

$$P(A) = \frac{L(A)}{L(S)} = \frac{d - c}{b - a}$$

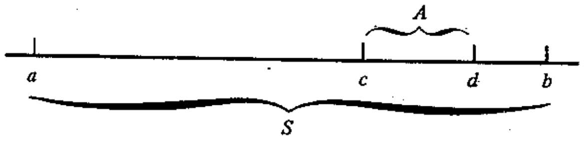

Figure 2.3.

if A is any continuous subinterval of S. For any subset B of S which is a union of nonoverlapping continuous subintervals of S, we define $L(B)$ to be the sum of the lengths of the nonoverlapping subintervals belonging to B. That is, if

$$B = C_1 \cup C_2 \cup \cdots, \quad \text{where } C_i \cap C_j = \varnothing, \quad \text{for } i \neq j,$$

then

$$L(B) = L(C_1) + L(C_2) + \cdots$$

and we have

$$P(B) = \frac{L(B)}{L(S)} = \frac{L(C_1)}{L(S)} + \frac{L(C_2)}{L(S)} + \cdots = P(C_1) + P(C_2) + \cdots.$$

Theorem 2.8.1. The rule

$$P(A) = \frac{L(A)}{L(S)} \quad \text{for } A \subset S$$

satisfies the three axioms.

Proof: $P(S) = L(S)/L(S) = 1$ so axiom 1 is satisfied. $L(A) \geq 0$ for any $A \subset S$ so $P(A) = L(A)/L(S) \geq 0$ and axiom 2 is satisfied. By the way in which lengths of unions of intervals are defined, we have

$$P(C_1 \cup C_2 \cup \cdots) = \frac{L(C_1 \cup C_2 \cdots)}{L(S)}$$

$$= \frac{L(C_1)}{L(S)} + \frac{L(C_2)}{L(S)} + \cdots = P(C_1) + P(C_2) + \cdots$$

if $C_i \cap C_j = \varnothing$ for all $i \neq j$. ◀

Note that if A is a single-element event then $L(A) = 0$ and we have $P(A) = 0$, as discussed above.

Let us now consider some specific examples.

Example 2.8.3. Doug is a 2-year-old boy. From his family history it seems plausible to assume that his adult height is equally likely to lie between 5 feet 9 inches and 6 feet 2 inches. Making this assumption, what is the probability that he will stand at least 6 feet high as an adult? What is the probability that his adult height will lie between 5 feet 10 inches and 5 feet 11 inches?

We use as our sample space

$$S = \{x\colon\ 69 \leq x \leq 74\}$$

where we are recording his achieved adult height in inches. Define

$$A = \{x\colon\ 72 \leq x \leq 74\}$$
$$B = \{x\colon\ 70 \leq x \leq 71\}.$$

Then

$$L(S) = 5, \qquad L(A) = 2, \qquad L(B) = 1$$

and we have

$$P(A) = \tfrac{2}{5}, \qquad P(B) = \tfrac{1}{5}.$$

Example 2.8.4. Assume that you daily ride a commuter train from your home in Connecticut into Manhattan. The station you leave from has trains leaving for Manhattan at 7 a.m., 7:13 a.m., 7:20 a.m., 7:25 a.m., 7:32 a.m., 7:45 a.m., and 7:55 a.m. It is your practice to take the first train that leaves after your arrival at the station. Due to the vagaries of your rising time and the traffic you encounter driving to the station, you are equally likely to arrive at the station at any instant between 7:15 a.m. and 7:45 a.m. On a particular day, what is the probability that you have to wait less than 5 minutes at the station? Less than 10 minutes? Suppose the 7:25 a.m. and 7:45 a.m. trains are expresses. What is the probability that you catch an express on a given day?

Let us, for convenience, take as our sample space

$$S = \{x\colon\ 0 \leq x \leq 30\}$$

where the elements of S are actually meant to represent minutes after 7:15 a.m. that might occur until your arrival time. (Refer to Figure 2.4.) Define the events

A: you wait less than 5 minutes
B: you wait less than 10 minutes
C: you catch an express.

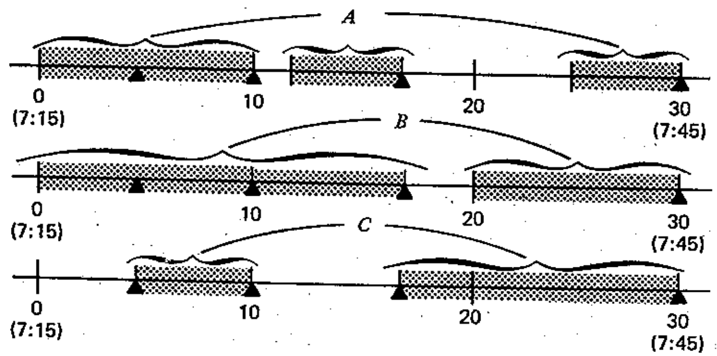

Figure 2.4.

Then

$$A = \{x: 0 \le x < 10 \text{ or } 12 \le x < 17 \text{ or } 25 \le x < 30\}$$
$$B = \{x: 0 \le x < 17 \text{ or } 20 \le x < 30\}$$
$$C = \{x: 5 \le x < 10 \text{ or } 17 \le x < 30\}$$

and

$$L(S) = 30, \quad L(A) = 20, \quad L(B) = 27, \quad L(C) = 18.$$

Thus

$$P(A) = \tfrac{2}{3}, \quad P(B) = \tfrac{9}{10}, \quad P(C) = \tfrac{3}{5}.$$

EXERCISE 2.8.

1. A fair die is rolled until a 1 occurs. Compute the probability that:
(a) 10 rolls are needed
(b) less than 4 rolls are needed
(c) an odd number of rolls is needed.

2. A fair pair of dice is rolled until a 7 occurs (as the sum of the 2 numbers on the dice). Compute the probability that
(a) 2 rolls are needed
(b) an even number of rolls is needed.

3. You fire a rifle at a target until you hit it. Assume the probability that you hit it is .9 for each shot and that the shots are independent. Compute the probability that:
(a) it takes more than 2 shots.
(b) the number of shots required is a multiple of 3.

4. Hugh takes a written driver's license test repeatedly until he passes it. Assume the probability that he passes it any given time is .1 and that the tests are independent. Compute the probability that:
(a) it takes him more than 4 attempts
(b) it takes him more than 10 attempts.

5. A traffic light on a route you travel every day turns red every 4 minutes, stays red 1 minute and then turns green again (thus it is green 3 minutes, red 1, etc.), with the red part of the signal starting on the hour, every hour.
(a) If you arrive at the light at a random instant between 7:55 a.m. and 8:05 a.m., what is the probability that you have to stop at the light?
(b) If you arrive at the light at a random instant between 7:54 a.m. and 8:04 a.m. what is the probability that you have to stop for the light?

6. The plug on an electric clock with a sweep second hand is pulled at a random instant of time within a certain minute. What is the probability that the second hand is between the 4 and the 5? Between the 1 and the 2? Between the 1 and the 6?

7. A point is chosen at random between 0 and 1 on the x-axis in the (x, y) plane. A circle centered at the origin is then drawn in the plane, with radius determined by the chosen point. Compute the probability that the area of the circle is less than $\pi/2$.

8. A 12-inch ruler is broken into 2 pieces at a random point along its length. What is the probability that the longer piece is at least twice the length of the shorter piece?