Chapter 2

Probability

2.1 Sample Space

In the study of statistics, we are concerned basically with the presentation and interpretation of **chance outcomes** that occur in a planned study or scientific investigation. For example, we may record the number of accidents that occur monthly at the intersection of Driftwood Lane and Royal Oak Drive, hoping to justify the installation of a traffic light; we might classify items coming off an assembly line as “defective” or “nondefective”; or we may be interested in the volume of gas released in a chemical reaction when the concentration of an acid is varied. Hence, the statistician is often dealing with either numerical data, representing counts or measurements, or **categorical data**, which can be classified according to some criterion.

We shall refer to any recording of information, whether it be numerical or categorical, as an **observation**. Thus, the numbers 2, 0, 1, and 2, representing the number of accidents that occurred for each month from January through April during the past year at the intersection of Driftwood Lane and Royal Oak Drive, constitute a set of observations. Similarly, the categorical data $N, D, N, N,$ and D, representing the items found to be defective or nondefective when five items are inspected, are recorded as observations.

Statisticians use the word **experiment** to describe any process that generates a set of data. A simple example of a statistical experiment is the tossing of a coin. In this experiment, there are only two possible outcomes, heads or tails. Another experiment might be the launching of a missile and observing of its velocity at specified times. The opinions of voters concerning a new sales tax can also be considered as observations of an experiment. We are particularly interested in the observations obtained by repeating the experiment several times. In most cases, the outcomes will depend on chance and, therefore, cannot be predicted with certainty. If a chemist runs an analysis several times under the same conditions, he or she will obtain different measurements, indicating an element of chance in the experimental procedure. Even when a coin is tossed repeatedly, we cannot be certain that a given toss will result in a head. However, we know the entire set of possibilities for each toss.

Given the discussion in Section 1.7, we should deal with the breadth of the term *experiment*. Three types of statistical studies were reviewed, and several examples were given of each. In each of the three cases, *designed experiments*, *observational studies*, and *retrospective studies*, the end result was a set of *data* that of course is subject to **uncertainty**. Though only one of these has the word *experiment* in its description, the process of generating the data or the process of observing the data is part of an experiment. The corrosion study discussed in Section 1.2 certainly involves an experiment, with measures of corrosion representing the data. The example given in Section 1.7 in which blood cholesterol and sodium were observed on a group of individuals represented an observational study (as opposed to a *designed* experiment), and yet the process generated data and the outcome is subject to uncertainty. Thus, it is an experiment. A third example in Section 1.7 represented a retrospective study in which historical data on monthly electric power consumption and average monthly ambient temperature were observed. Even though the data may have been in the files for decades, the process is still referred to as an experiment.

Definition 2.1: The set of all possible outcomes of a statistical experiment is called the **sample space** and is represented by the symbol S.

Each outcome in a sample space is called an **element** or a **member** of the sample space, or simply a **sample point**. If the sample space has a finite number of elements, we may *list* the members separated by commas and enclosed in braces. Thus, the sample space S, of possible outcomes when a coin is flipped, may be written

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

where H and T correspond to heads and tails, respectively.

Example 2.1: Consider the experiment of tossing a die. If we are interested in the number that shows on the top face, the sample space is

$$S_1 = \{1, 2, 3, 4, 5, 6\}.$$

If we are interested only in whether the number is even or odd, the sample space is simply

$$S_2 = \{\text{even, odd}\}.$$

Example 2.1 illustrates the fact that more than one sample space can be used to describe the outcomes of an experiment. In this case, S_1 provides more information than S_2. If we know which element in S_1 occurs, we can tell which outcome in S_2 occurs; however, a knowledge of what happens in S_2 is of little help in determining which element in S_1 occurs. In general, it is desirable to use the sample space that gives the most information concerning the outcomes of the experiment. In some experiments, it is helpful to list the elements of the sample space systematically by means of a **tree diagram**.

Example 2.2: An experiment consists of flipping a coin and then flipping it a second time if a head occurs. If a tail occurs on the first flip, then a die is tossed once. To list the elements of the sample space providing the most information, we construct the tree diagram of Figure 2.1. The various paths along the branches of the tree give the distinct sample points. Starting with the top left branch and moving to the right along the first path, we get the sample point HH, indicating the possibility that heads occurs on two successive flips of the coin. Likewise, the sample point $T3$ indicates the possibility that the coin will show a tail followed by a 3 on the toss of the die. By proceeding along all paths, we see that the sample space is

$$S = \{HH,\ HT,\ T1,\ T2,\ T3,\ T4,\ T5,\ T6\}.$$

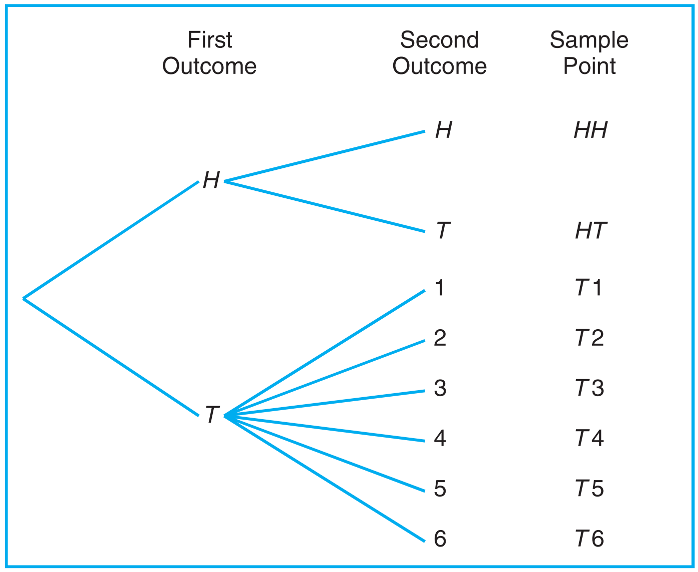

Figure 2.1: Tree diagram for Example 2.2.

Many of the concepts in this chapter are best illustrated with examples involving the use of dice and cards. These are particularly important applications to use early in the learning process, to facilitate the flow of these new concepts into scientific and engineering examples such as the following.

Example 2.3: Suppose that three items are selected at random from a manufacturing process. Each item is inspected and classified defective, D, or nondefective, N. To list the elements of the sample space providing the most information, we construct the tree diagram of Figure 2.2. Now, the various paths along the branches of the tree give the distinct sample points. Starting with the first path, we get the sample point DDD, indicating the possibility that all three items inspected are defective. As we proceed along the other paths, we see that the sample space is

$$S = \{DDD,\ DDN,\ DND,\ DNN,\ NDD,\ NDN,\ NND,\ NNN\}.$$

Sample spaces with a large or infinite number of sample points are best described by a **statement** or **rule method**. For example, if the possible outcomes of an experiment are the set of cities in the world with a population over 1 million, our sample space is written

$$S = \{x \mid x \text{ is a city with a population over 1 million}\},$$

which reads "S is the set of all x such that x is a city with a population over 1 million." The vertical bar is read "such that." Similarly, if S is the set of all points (x, y) on the boundary or the interior of a circle of radius 2 with center at the origin, we write the **rule**

$$S = \{(x, y) \mid x^2 + y^2 \leq 4\}.$$

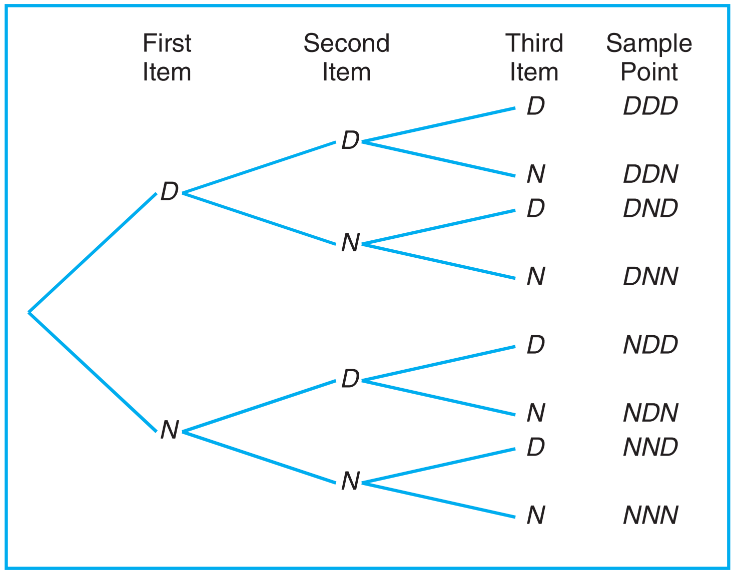

Figure 2.2: Tree diagram for Example 2.3.

Whether we describe the sample space by the rule method or by listing the elements will depend on the specific problem at hand. The rule method has practical advantages, particularly for many experiments where listing becomes a tedious chore.

Consider the situation of Example 2.3 in which items from a manufacturing process are either D, defective, or N, nondefective. There are many important statistical procedures called sampling plans that determine whether or not a "lot" of items is considered satisfactory. One such plan involves sampling until k defectives are observed. Suppose the experiment is to sample items randomly until one defective item is observed. The sample space for this case is

$$S = \{D, ND, NND, NNND, \ldots\}.$$

2.2 Events

For any given experiment, we may be interested in the occurrence of certain **events** rather than in the occurrence of a specific element in the sample space. For instance, we may be interested in the event A that the outcome when a die is tossed is divisible by 3. This will occur if the outcome is an element of the subset $A = \{3, 6\}$ of the sample space S_1 in Example 2.1. As a further illustration, we may be interested in the event B that the number of defectives is greater than 1 in Example 2.3. This will occur if the outcome is an element of the subset

$$B = \{DDN, DND, NDD, DDD\}$$

of the sample space S.

To each event we assign a collection of sample points, which constitute a subset of the sample space. That subset represents all of the elements for which the event is true.

Definition 2.2: An **event** is a subset of a sample space.

Example 2.4: Given the sample space $S = \{t \mid t \geq 0\}$, where t is the life in years of a certain electronic component, then the event A that the component fails before the end of the fifth year is the subset $A = \{t \mid 0 \leq t < 5\}$.

It is conceivable that an event may be a subset that includes the entire sample space S or a subset of S called the **null set** and denoted by the symbol ϕ, which contains no elements at all. For instance, if we let A be the event of detecting a microscopic organism by the naked eye in a biological experiment, then $A = \phi$. Also, if

$$B = \{x \mid x \text{ is an even factor of } 7\},$$

then B must be the null set, since the only possible factors of 7 are the odd numbers 1 and 7.

Consider an experiment where the smoking habits of the employees of a manufacturing firm are recorded. A possible sample space might classify an individual as a nonsmoker, a light smoker, a moderate smoker, or a heavy smoker. Let the subset of smokers be some event. Then all the nonsmokers correspond to a different event, also a subset of S, which is called the **complement** of the set of smokers.

Definition 2.3: The **complement** of an event A with respect to S is the subset of all elements of S that are not in A. We denote the complement of A by the symbol A'.

Example 2.5: Let R be the event that a red card is selected from an ordinary deck of 52 playing cards, and let S be the entire deck. Then R' is the event that the card selected from the deck is not a red card but a black card.

Example 2.6: Consider the sample space

$$S = \{\text{book, cell phone, mp3, paper, stationery, laptop}\}.$$

Let $A = \{\text{book, stationery, laptop, paper}\}$. Then the complement of A is $A' = \{\text{cell phone, mp3}\}$.

We now consider certain operations with events that will result in the formation of new events. These new events will be subsets of the same sample space as the given events. Suppose that A and B are two events associated with an experiment. In other words, A and B are subsets of the same sample space S. For example, in the tossing of a die we might let A be the event that an even number occurs and B the event that a number greater than 3 shows. Then the subsets $A = \{2, 4, 6\}$ and $B = \{4, 5, 6\}$ are subsets of the same sample space

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

Note that *both* A and B will occur on a given toss if the outcome is an element of the subset $\{4, 6\}$, which is just the **intersection** of A and B.

Definition 2.4: The **intersection** of two events A and B, denoted by the symbol $A \cap B$, is the event containing all elements that are common to A and B.

Example 2.7: Let E be the event that a person selected at random in a classroom is majoring in engineering, and let F be the event that the person is female. Then $E \cap F$ is the event of all female engineering students in the classroom.

Example 2.8: Let $V = \{a, e, i, o, u\}$ and $C = \{l, r, s, t\}$; then it follows that $V \cap C = \phi$. That is, V and C have no elements in common and, therefore, cannot both simultaneously occur.

For certain statistical experiments it is by no means unusual to define two events, A and B, that cannot both occur simultaneously. The events A and B are then said to be **mutually exclusive**. Stated more formally, we have the following definition:

Definition 2.5: Two events A and B are **mutually exclusive**, or **disjoint**, if $A \cap B = \phi$, that is, if A and B have no elements in common.

Example 2.9: A cable television company offers programs on eight different channels, three of which are affiliated with ABC, two with NBC, and one with CBS. The other two are an educational channel and the ESPN sports channel. Suppose that a person subscribing to this service turns on a television set without first selecting the channel. Let A be the event that the program belongs to the NBC network and B the event that it belongs to the CBS network. Since a television program cannot belong to more than one network, the events A and B have no programs in common. Therefore, the intersection $A \cap B$ contains no programs, and consequently the events A and B are mutually exclusive.

Often one is interested in the occurrence of at least one of two events associated with an experiment. Thus, in the die-tossing experiment, if

$$A = \{2, 4, 6\} \text{ and } B = \{4, 5, 6\},$$

we might be interested in either A or B occurring or both A and B occurring. Such an event, called the **union** of A and B, will occur if the outcome is an element of the subset $\{2, 4, 5, 6\}$.

Definition 2.6: The **union** of the two events A and B, denoted by the symbol $A \cup B$, is the event containing all the elements that belong to A or B or both.

Example 2.10: Let $A = \{a, b, c\}$ and $B = \{b, c, d, e\}$; then $A \cup B = \{a, b, c, d, e\}$.

Example 2.11: Let P be the event that an employee selected at random from an oil drilling company smokes cigarettes. Let Q be the event that the employee selected drinks alcoholic beverages. Then the event $P \cup Q$ is the set of all employees who either drink or smoke or do both.

Example 2.12: If $M = \{x \mid 3 < x < 9\}$ and $N = \{y \mid 5 < y < 12\}$, then

$$M \cup N = \{z \mid 3 < z < 12\}.$$

The relationship between events and the corresponding sample space can be illustrated graphically by means of **Venn diagrams**. In a Venn diagram we let the sample space be a rectangle and represent events by circles drawn inside the rectangle. Thus, in Figure 2.3, we see that

$$A \cap B = \text{ regions 1 and 2,}$$
$$B \cap C = \text{ regions 1 and 3,}$$

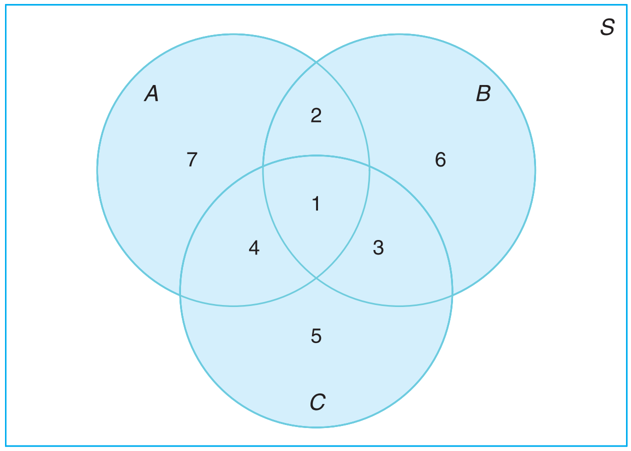

Figure 2.3: Events represented by various regions.

$$
\begin{aligned}
A \cup C &= \text{ regions 1, 2, 3, 4, 5, and 7,}\\
B' \cap A &= \text{ regions 4 and 7,}\\
A \cap B \cap C &= \text{ region 1,}\\
(A \cup B) \cap C' &= \text{ regions 2, 6, and 7,}
\end{aligned}
$$

and so forth.

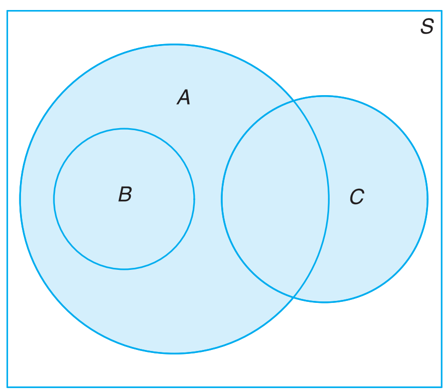

Figure 2.4: Events of the sample space S.

In Figure 2.4, we see that events A, B, and C are all subsets of the sample space S. It is also clear that event B is a subset of event A; event $B \cap C$ has no elements and hence B and C are mutually exclusive; event $A \cap C$ has at least one element; and event $A \cup B = A$. Figure 2.4 might, therefore, depict a situation where we select a card at random from an ordinary deck of 52 playing cards and observe whether the following events occur:

A: the card is red,

B: the card is the jack, queen, or king of diamonds,

C: the card is an ace.

Clearly, the event $A \cap C$ consists of only the two red aces.

Several results that follow from the foregoing definitions, which may easily be verified by means of Venn diagrams, are as follows:

1. $A \cap \phi = \phi$.
2. $A \cup \phi = A$.
3. $A \cap A' = \phi$.
4. $A \cup A' = S$.
5. $S' = \phi$.
6. $\phi' = S$.
7. $(A')' = A$.
8. $(A \cap B)' = A' \cup B'$.
9. $(A \cup B)' = A' \cap B'$.

Exercises

2.1 List the elements of each of the following sample spaces:

(a) the set of integers between 1 and 50 divisible by 8;

(b) the set $S = \{x \mid x^2 + 4x - 5 = 0\}$;

(c) the set of outcomes when a coin is tossed until a tail or three heads appear;

(d) the set $S = \{x \mid x \text{ is a continent}\}$;

(e) the set $S = \{x \mid 2x - 4 \geq 0 \text{ and } x < 1\}$.

2.2 Use the rule method to describe the sample space S consisting of all points in the first quadrant inside a circle of radius 3 with center at the origin.

2.3 Which of the following events are equal?

(a) $A = \{1, 3\}$;

(b) $B = \{x \mid x \text{ is a number on a die}\}$;

(c) $C = \{x \mid x^2 - 4x + 3 = 0\}$;

(d) $D = \{x \mid x \text{ is the number of heads when six coins are tossed}\}$.

2.4 An experiment involves tossing a pair of dice, one green and one red, and recording the numbers that come up. If x equals the outcome on the green die and y the outcome on the red die, describe the sample space S

(a) by listing the elements (x, y);

(b) by using the rule method.

2.5 An experiment consists of tossing a die and then flipping a coin once if the number on the die is even. If the number on the die is odd, the coin is flipped twice. Using the notation $4H$, for example, to denote the outcome that the die comes up 4 and then the coin comes up heads, and $3HT$ to denote the outcome that the die comes up 3 followed by a head and then a tail on the coin, construct a tree diagram to show the 18 elements of the sample space S.

2.6 Two jurors are selected from 4 alternates to serve at a murder trial. Using the notation A_1A_3, for example, to denote the simple event that alternates 1 and 3 are selected, list the 6 elements of the sample space S.

2.7 Four students are selected at random from a chemistry class and classified as male or female. List the elements of the sample space S_1, using the letter M for male and F for female. Define a second sample space S_2 where the elements represent the number of females selected.

2.8 For the sample space of Exercise 2.4,

(a) list the elements corresponding to the event A that the sum is greater than 8;

(b) list the elements corresponding to the event B that a 2 occurs on either die;

(c) list the elements corresponding to the event C that a number greater than 4 comes up on the green die;

(d) list the elements corresponding to the event $A \cap C$;

(e) list the elements corresponding to the event $A \cap B$;

(f) list the elements corresponding to the event $B \cap C$;

(g) construct a Venn diagram to illustrate the intersections and unions of the events A, B, and C.

2.9 For the sample space of Exercise 2.5,

(a) list the elements corresponding to the event A that a number less than 3 occurs on the die;

(b) list the elements corresponding to the event B that two tails occur;

(c) list the elements corresponding to the event A';