CHAPTER 2

Probability

1 INTRODUCTION

An individual's approach to probability depends on the nature of his interest in the subject. The pure mathematician usually prefers to treat probability from an axiomatic point of view, just as he does, say, the study of geometry. The applied statistician usually prefers to think of probability as the proportion of times that a certain event will occur if the experiment related to the event is repeated indefinitely. The approach to probability here is based on a blending of these two points of view.

The statistician is usually interested in probability only as it pertains to the possible outcomes of experiments. Furthermore, most statisticians are interested in only those experiments that are repetitive in nature or that can be conceived of as being so. Experiments such as tossing a coin, counting the number of defective parts in a box of parts, or reading the daily temperature on a thermometer are examples of simple repetitive experiments. An experiment in which several experimental animals are fed different rations may be performed only once with those same animals; nevertheless, the experiment may be thought of as the first in an unlimited number of similar experiments and therefore may be conceived of as being repetitive.

2 SAMPLE SPACE

Consider a simple experiment such as tossing a coin. In this experiment there are but two possible outcomes, a head and a tail. It is convenient to represent the possible outcomes of such an experiment, and experiments in general, by points on a line or by points in higher dimensions. Here it would be convenient to represent a head by the point 1 on the x axis and a tail by the point 0. This choice is convenient because the number corresponds to the number of heads obtained in the toss. If the experiment had consisted of

2.2

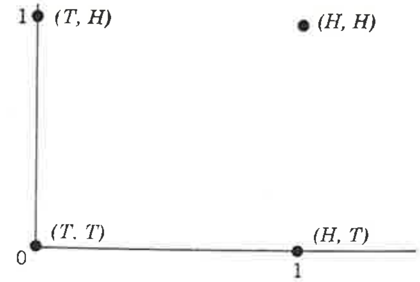

Fig. 1. A simple sample space.

tossing the coin twice, there would have been four possible outcomes, namely HH, HT, TH, and TT. For reasons of symmetry, it would be desirable to represent these outcomes by the points (1, 1), (1, 0), (0, 1), and (0, 0) in the x, y plane. Figure 1 illustrates this choice of points to represent the possible outcomes of the experiment.

If the coin were tossed three times, it would be convenient to use three dimensions to represent the possible experimental outcomes. This representation, of course, is merely a convenience, and if desired one could just as well mark off any eight points on the x axis to represent the eight possible outcomes.

In the experiment of rolling two dice, there are 36 possible outcomes, which have been listed in Table 1. The first number of each pair denotes

TABLE 1

11	21	31	41	51	61
12	22	32	42	52	62
13	23	33	43	53	63
14	24	34	44	54	64
15	25	35	45	55	65
16	26	36	46	56	66

the number that came up on one of the dice and the second number denotes the number that came up on the other die. It is assumed that the two dice are distinguishable or are rolled in order. For this experiment a natural set of points to represent the possible outcomes are the 36 points in the x, y plane whose coordinates are the corresponding number pairs of Table 1. This choice is shown in Fig. 2.

An experiment that consists of reading the temperature of a patient in a hospital has a very large number of possible outcomes depending on the degree of accuracy with which the thermometer is read. For such an experiment it is convenient to assume that the patient's temperature can assume

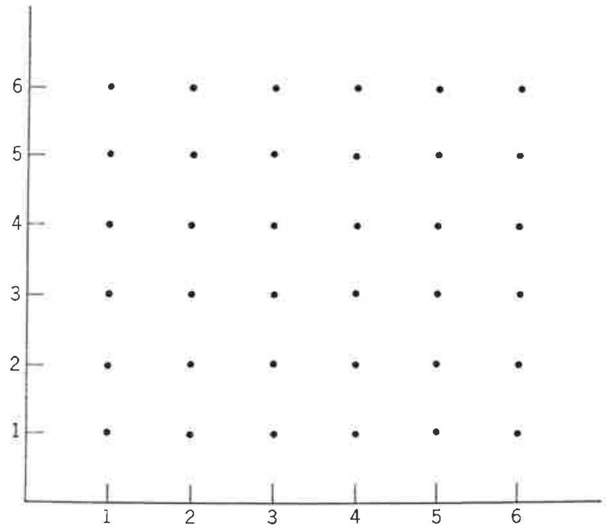

Fig. 2. A sample space for rolling two dice.

any value between, say, 95° and 110°; therefore the possible outcomes would be represented in a natural way by the points inside the interval from 95 to 110 on the x axis. This, of course, is a convenient idealization and ignores the impossibility of reading a thermometer to unlimited accuracy.

DEFINITION: *The set of points representing the possible outcomes of an experiment is called the sample space of the experiment.*

The idea of a sample space is introduced because it is a convenient mathematical device for developing the theory of probability as it pertains to the outcomes of experiments.

3 EVENTS

Consider an experiment such that whatever the outcome of the experiment it can be decided whether an event A has occurred. This means that each sample point can be classified as one for which A will occur or as one for which A will not occur. Thus, if A is the event of getting exactly one head and one tail in tossing a coin twice, the two sample points (H, T) and (T, H) of Fig. 1 correspond to the occurrence of A. If A is the event of getting a total of seven points in rolling two dice, then A is associated with the six sample points (1, 6), (2, 5), (3, 4), (4, 3), (5, 2), and (6, 1) of Fig. 2. If A is the event that a patient's temperature will be at least as high as 102, then A will consist of the interval of points from 102 to 110 on the x axis.

2.3

DEFINITION: *An event is a subset of a sample space.*

Since a subset of a set of points is understood to include the possibility that the subset is the entire set of points or that it contains none of the points of the set, this definition includes an event that is certain to occur or one that cannot possibly occur when the experiment is performed.

In view of the correspondence between events and sets of points the study of the relationship between various events is reduced to the study of the relationship between the corresponding sets. For this purpose it is convenient to represent the sample space, whatever its dimension or whatever the number of points in it, by a set of points inside a rectangle in a plane. An event A, which is therefore a subset of the points in this rectangle, is represented by the points lying inside a closed curve contained in the rectangle. If B is some other event of interest, it will be represented by the points inside some other closed curve in the rectangle. This representation is shown in Fig. 3. No attempt has been made to indicate whether the number of points is finite or infinite because a representation for both types of sample spaces is desired.

If A and B are two events associated with an experiment, one may be interested in knowing whether at least one of the events A and B will occur when the experiment is performed. Now the set of points that consists of all points that belong to A, or B, or both A and B is called the union of A and B and is denoted by the symbol $A \cup B$. This set of points, which is shown as the shaded region in Fig. 3, therefore represents the event that at least one of the events A and B will occur.

As an illustration, if A is the event of getting a six on the first die when rolling two dice and B is the event of getting a six on the second die, then $A \cup B$ is the event of getting at least one six in rolling two dice. The event A consists of the six points found in the last column of points in the sample space shown in Fig. 2 and the event B consists of the six points found in the last row of points in that sample space. The event $A \cup B$ is then the set of eleven points found in the union of the last column and last row of points.

Another event of possible interest is that of knowing whether both events A and B will occur when the experiment is performed. The set of points that consists of all points that belong to both A and B is called the intersection of

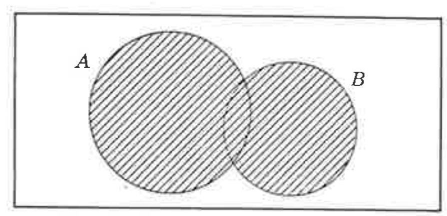

Fig. 3. Representation of $A \cup B$.

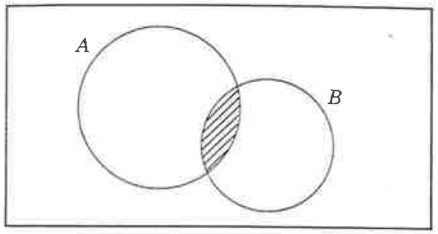

Fig. 4. Representation of $A \cap B$.

A and B and is denoted by the symbol $A \cap B$. This set of points, which is shown as the shaded region in Fig. 4, therefore represents the event that both A and B will occur when the experiment is performed.

In the preceding illustration concerning the two dice, $A \cap B$ is the event that both dice will show a six. It is represented by the single point (6, 6), which is the intersection of the last column, and last row, sets of points.

Corresponding to any event A there is an associated event, denoted by $\bar{A}$, which states that A will not occur when the experiment is performed. It is represented by all the points of the rectangle not found in A and is shown as the shaded region in Fig. 5. The set $\bar{A}$ is called the complement of the set A relative to the sample space.

If two sets, A and B, have no points in common they are said to be disjoint sets. In the language of events, such events are called *disjoint events*, but they are also called *mutually exclusive events*, because the occurrence of one of those events excludes the possible occurrence of the other.

4 PROBABILITY

The familiar functions of calculus are what are known as point functions. The function defined by the formula $f(x) = x^2$ is an example of a point function, for to each point on the x axis this formula assigns the value of the function. The notion of function is much broader than this, however, and permits the elements of the domain of the function to be sets of points rather than individual points. In this case the function is called a *set function.* As

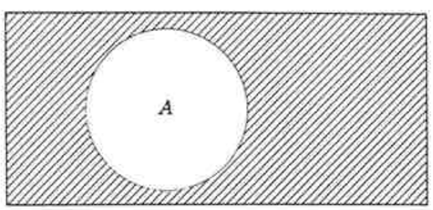

Fig. 5. Representation of $\bar{A}$.

2.4

an illustration, the domain might consist of all intervals on the x axis and the function might be the function that gives the length of the interval. Set functions are introduced here because they are needed for defining probability.

Probability is an idealization of the proportion of times that a certain result will occur in repeated trials of an experiment; therefore a probability model for events should be one for which the probability that an event A will occur, which is denoted by $P\{A\}$, is equal to the proportion of times that the event A would be expected to occur in repeated trials of the experiment. Since $P\{A\}$ is a function defined on sets, such as A, it is a set function.

If an experiment could be repeated a large number of times the results could be used to assign a value to $P\{A\}$. However, there is no requirement that the experiment be performed at all before such a probability is assigned. Thus, if A is the event of getting two sixes in rolling two dice, symmetry considerations would suggest the value $\frac{1}{36}$ for the probability that this event will occur. The experimenter is free to assign any value he desires, but if his assignment is unrealistic from the long run proportion point of view his probability model is not likely to prove very useful to him for making predictions about future experiments.

If probabilities of events are to be interpreted as models for the proportion of times those events will occur in repetitions of the experiment, such probabilities should possess the essential properties of proportions. Thus, a probability should be a number between 0 and 1 because a proportion is that kind of number. Further, the probability of the event S, where S is the sample space, should be 1 because some one of the possible outcomes is certain to occur when the experiment is performed. Finally, if two events A and B are disjoint, the probability of the union of those events should be equal to the sum of the probabilities of the two events because for such events the proportion of times that A or B occurs will be equal to the proportion of times that A occurs plus the proportion of times that B occurs. It has been found that any other reasonable properties of probabilities that might be desired will be satisfied if the following three conditions, which are based on the preceding discussion, are satisfied. These are called the axioms, or postulates, of probability. They place restrictions on the type of set function P that can be used to calculate probabilities of events. Such a function is called a probability measure.

AXIOMS OF PROBABILITY: *A probability measure P is a real valued set function defined on a sample space S that satisfies*

(1) $0 \leq P\{A\} \leq 1$ *for every event A*

(2) $P\{S\} = 1$

(3) $P\{A_1 \cup A_2 \cup \cdots\} = P\{A_1\} + P\{A_2\} + \cdots$ *for every finite or infinite sequence of disjoint events $A_1, A_2, \cdots$.*