Chapter Seven

Probability Theory

7.1 INTRODUCTION

Probability theory is a mathematical modeling of the phenomenon of chance or randomness. If a coin is tossed in a random manner, it can land heads or tails, but we do not know which of these will occur in a single toss. However, suppose we let s be the number of times heads appears when the coin is tossed n times. As n increases, the ratio $f = s/n$, called the *relative frequency* of the outcome, becomes more stable. If the coin is perfectly balanced, then we expect that the coin will land heads approximately 50% of the time or, in other words, the relative frequency will approach $\frac{1}{2}$. Alternatively, assuming the coin is perfectly balanced, we can arrive at the value $\frac{1}{2}$ deductively. That is, any side of the coin is as likely to occur as the other; hence the chance of getting a head is 1 in 2 which means the probability of getting a head is $\frac{1}{2}$. Although the specific outcome on any one toss is unknown, the behavior over the long run is determined. This stable long-run behavior of random phenomena forms the basis of probability theory.

A probabilistic mathematical model of random phenomena is defined by assigning "probabilities" to all the possible outcomes of an experiment. The reliability of our mathematical model for a given experiment depends upon the closeness of the assigned probabilities to the actual limiting relative frequencies. This then gives rise to problems of testing and reliability, which form the subject matter of statistics and which lie beyond the scope of this text.

7.2 SAMPLE SPACE AND EVENTS

The set S of all possible outcomes of a given experiment is called the *sample space*. A particular outcome, i.e., an element in S, is called a *sample point*. An *event* A is a set of outcomes or, in other words, a subset of the sample space S. In particular, the set $\{a\}$ consisting of a single sample point $a \in S$ is called an *elementary event*. Furthermore, the empty set $\varnothing$ and S itself are subsets of S and so are events; $\varnothing$ is sometimes called the *impossible event* or the *null event*.

Since an event is a set, we can combine events to form new events using the various set operations:

(i) $A \cup B$ is the event that occurs iff A occurs or B occurs (or both).

(ii) $A \cup B$ is the event that occurs iff A occurs and B occurs.

(iii) A^c, the complement of A, also written $\overline{A}$, is the event that occurs iff A does *not* occur.

Two events A and B are called *mutually exclusive* if they are disjoint, that is, if $A \cap B = \varnothing$. In other words, A and B are mutually exclusive iff they cannot occur simultaneously. Three or more events are mutually exclusive if every two of them are mutually exclusive.

Example 7.1

(a) **Experiment:** Toss a die and observe the number (of dots) that appears on top.
The sample space S consists of the six possible numbers; that is,

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

Let A be the event that an even number occurs, B that an odd number occurs, and C that a prime number occurs; that is, let

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

Then
$A \cup C = \{2, 3, 4, 5, 6\}$ is the event that an even or a prime number occurs.
$B \cap C = \{3, 5\}$ is the event that an odd prime number occurs.
$C^c = \{1, 4, 6\}$ is the event that a prime number does not occur.

Note that A and B are mutually exclusive: $A \cap B = \varnothing$. In other words, an even number and an odd number cannot occur simultaneously.

(b) **Experiment:** Toss a coin three times and observe the sequence of heads (T) and tails (T) that appears. The sample space S consists of the following eight elements:

$$S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$

Let A be the event that two or more heads appear consecutively, and B that all the tosses are the same; that is, let

$$A = \{HHH, HHT, THH\}; \qquad \text{and} \qquad B = \{HHH, TTT\}$$

Then $A \cap B = \{HHH\}$ is the elementary event in which only heads appear. The event that five heads appear is the empty set $\varnothing$.

(c) **Experiment:** Toss a coin until a head appears, and then count the number of times the coin is tossed.
The sample space of this experiment is $S = \{1, 2, 3, \ldots\}$. Since every positive integer is an element of S, the sample space is infinite.

Remark: The sample space S in Example 7.1(c), as noted, is not finite. The theory concerning such samples space lies beyond the scope of this text. Thus, unless otherwise stated, all our sample spaces S shall be finite.

7.3 FINITE PROBABILITY SPACES

The following definition applies.

Definition: Let S be a finite sample space, say $S = \{a_1, a_2, ..., a_n\}$. A *finite probability space*, or *probability model*, is obtained by assigning to each point a_i in S a real number p_i, called the *probability* of a_i, satisfying the following properties:

(i) Each p_i is nonnegative, that is, $p_i \geq 0$.
(ii) The sum of the p_i is 1, that is, $p_1 + p_2 + \cdots + p_n = 1$.

The *probability* of an event A written $P(A)$, is then defined to be the sum of the probabilities of the points in A.

The singleton set $\{a_i\}$ is called an *elementary* event and, for notational convenience, we write $P(a_i)$ for $P(\{a_i\})$.

Example 7.2

Experiment: Let three coins be tossed and the number of heads observed. [Compare with the above Example 7.1(b).]

The sample space is $S = \{0, 1, 2, 3\}$. The following assignments on the elements of S defines a probability space:

$$P(0) = \frac{1}{8}, \quad P(1) = \frac{3}{8}, \quad P(2) = \frac{3}{8}, \quad P(3) = \frac{1}{8}$$

That is, each probability is nonnegative, and the sum of the probabilities is 1. Let A be the event that at least one head appears, and let B be the event that all heads or all tails appear; that is, let $A = \{1, 2, 3\}$ and $B = \{0, 3\}$. Then, by definition,

$$P(A) = P(1) + P(2) + P(3) = \frac{3}{8} + \frac{3}{8} + \frac{1}{8} = \frac{7}{8} \quad \text{and} \quad P(B) = P(0) + P(3) = \frac{1}{8} + \frac{1}{8} = \frac{1}{4}$$

Equiprobable Spaces

Frequently, the physical characteristics of an experiment suggest that the various outcomes of the sample space be assigned equal probabilities. Such a finite probability space S, where each sample point has the same probability, will be called an *equiprobable space*. In particular, if S contains n points, then the probability of each point is $1/n$. Furthermore, if an event A contains r points, then its probability is $r(1/n) = r/n$. In other words,

$$P(A) = \frac{\text{number of elements in } A}{\text{number of elements is } S} = \frac{n(A)}{n(S)} \quad \text{or} \quad P(A) = \frac{\text{number of outcomes favourable to } A}{\text{total number of possible outcomes}}$$

where $n(A)$ denotes the number of elements in a set A.

We emphasize that the above formula for $P(A)$ can only be used with respect to an equiprobable space, and cannot be used in general.

The expression *at random* will be used only with respect to an equiprobable space; the statement "choose a point at random from a set S" shall mean that every sample point in S has the same probability of being chosen.

Example 7.3

Let a card be selected from an ordinary deck of 52 playing cards. Let

$$A = \{\text{the card is a spade}\} \quad \text{and} \quad B = \{\text{the card is a face card}\}$$

(A face card is a jack, queen, or king.) We compute $P(A)$, $P(B)$, and $P(A \cap B)$. Since we have an equiprobable space,

$$P(A) = \frac{\text{number of spades}}{\text{number of cards}} = \frac{13}{52} = \frac{1}{4}, \qquad P(B) = \frac{\text{number of face cards}}{\text{number of cards}} = \frac{12}{52} = \frac{3}{13}$$

$$P(A \cap B) = \frac{\text{number of spade face cards}}{\text{number of cards}} = \frac{3}{32}$$

Theorems on Finite Probability Spaces

The following theorem follows directly from the fact that the probability of an event is the sum of the probabilities of its points.

Theorem 7.1: The probability function P defined on the class of all events in a finite probability space has the following properties:

[$\mathbf{P_1}$] For every event A, $0 \le P(A) \le 1$.

[$\mathbf{P_2}$] $P(S) = 1$.

[$\mathbf{P_3}$] If events A and B are mutually exclusive, then $P(A \cup B) = P(A) + P(B)$.

The next theorem formalizes our intuition that if p is the probability that an event E occurs, then $1 - p$ is the probability that E does not occur. (That is, if we hit a target $p = 1/3$ of the times, then we miss the target $1 - p = 2/3$ of the times.)

Theorem 7.2: Let A be any event. Then $P(A^c) = 1 - P(A)$.

The following theorem (proved in Problem 7.16) follows directly from Theorem 7.1.

Theorem 7.3: Let $\varnothing$ be the empty set, and suppose A and B are any events. Then:

(i) $P(\varnothing) = 0$.

(ii) $P(A \backslash B) = P(A) - P(A \cap B)$.

(iii) If $A \subseteq B$, then $P(A) \le P(B)$.

Observe that Property [$\mathbf{P}_3$] in Theorem 7.1 gives the probability of the union of events in the case that the events are disjoint. The general formula (proved in Problem 7.17) is called the Addition Principle. Specifically:

Theorem 7.4 (Addition Principle): For any events A and B,

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

Example 7.4

Suppose a student is selected at random from 100 students where 30 are taking mathematics, 20 are taking chemistry, and 10 taking mathematics and chemistry. Find the probability p that the student is taking mathematics or chemistry.

Let M = {students taking mathematics} and C = {students taking chemistry}. Since the space is equiprobable,

$$P(M) = \frac{30}{100} = \frac{3}{10} \qquad P(C) = \frac{20}{100} = \frac{1}{5}, \qquad P(M \text{ and } C) = P(M \cap C) = \frac{10}{100} = \frac{1}{10}$$

Thus, by the Addition Principle (Theorem 7.4),

$$p = P(M \text{ or } C) = P(M \cup C) = P(M) + P(C) - P(M \cap C) = \frac{3}{10} + \frac{1}{5} - \frac{1}{10} = \frac{2}{5}$$

7.4 CONDITIONAL PROBABILITY

Suppose E is an event in a sample space S with $P(E) > 0$. The probability that an event A occurs once E has occurred or, specifically, the *conditional probability of A given E, written $P(A|E)$*, is defined as follows:

$$P(A|E) = \frac{P(A \cap E)}{P(E)}$$

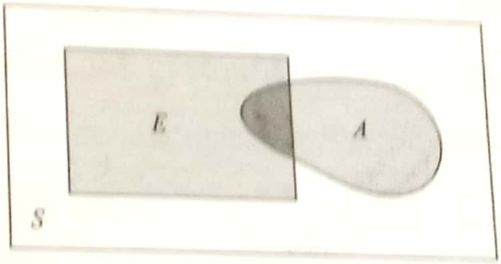

Fig. 7.1

As pictured in the Venn diagram in Fig. 7.1, $P(A|E)$ measures, in a certain sense, the relative probability of A with respect to the reduced space E.

Now suppose S is an equiprobable space, and we let $n(A)$ denote the number of elements in the event A. Then

$$P(A \cap E) = \frac{n(A \cap E)}{n(S)}, \quad P(E) = \frac{n(E)}{n(S)}, \quad \text{and so} \quad P(A|E) = \frac{P(A \cap E)}{P(E)} = \frac{n(A \cap E)}{n(E)}$$

We state this result formally.

Theorem 7.5: Suppose S is an equiprobable space and A and B are events. Then

$$P(A|E) = \frac{\text{number of elements in } A \cap E}{\text{number of elements in } E} = \frac{n(A \cap E)}{n(E)}$$

Example 7.5

(a) A pair of fair dice is tossed. The sample space S consists of the 36 ordered pairs (a, b) where a and b can be any of the integers from 1 to 6. (See Problem 7.3.) Thus the probability of any point is 1/36. Find the probability that one of the dice is 2 if the sum is 6. That is, find $P(A|E)$ where

$$E = \{\text{sum is } 6\} \quad \text{and} \quad A = \{2 \text{ appears on at least one die}\}$$

Also find $P(A)$.

Now E consists of five elements, specifically

$$E = \{(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)\}$$

Two of them, (2, 4) and (4, 2), belong to A; that is,

$$A \cap E = \{(2, 4), (4, 2)\}$$

By Theorem 7.5, $P(A|E) = 2/5$.

On the other hand, A consists of 11 elements, specifically,

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

and S consists of 36 elements. Hence $P(A) = 11/36$.

(b) A couple has two children; the sample space is $S = \{bb, bg, gb, gg\}$; with probability 1/4 for each point. Find the probability p that both children are boys if it is known that: (i) at least one of the children is a boy, (ii) the older child is a boy.

(i) Here the reduced space consists of three elements $\{bb, bg, gb\}$; hence $p = \frac{1}{3}$.

(ii) Here the reduced space consists of only two elements $\{bb, bg\}$; hence $p = \frac{1}{2}$.

Multiplication Theorem for Conditional Probability

Suppose A and B are events in a sample space S with $P(A) > 0$. By definition of conditional probability,

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

Multiplying both sides by $P(A)$ gives us the following useful result:

Theorem 7.6 (Multiplication Theorem for Conditional Probability):

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

The multiplication theorem gives us a formula for the probability that events A and B both occur. It can easily be extended to three or more events $A_1, A_2, \ldots A_m$; that is,

$$P(A_1 \cap A_2 \cap \cdots \cap A_m) = P(A_1) \cdot P(A_2|A_1) \cdots P(A_m|A_1 \cap A_2 \cap \cdots \cap A_{m-1})$$

Example 7.6

A lot contains 12 items of which four are defective. Three items are drawn at random from the lot one after the other. Find the probability p that all three are nondefective.

The probability that the first item is nondefective is $\frac{8}{12}$ since eight of 12 items are nondefective. If the first item is nondefective, then the probability that the next item is nondefective is $\frac{7}{11}$ since only seven of the remaining 11 items are nondefective. If the first two items are nondefective, then the probability that the last item is nondefective is $\frac{6}{10}$ since only 6 of the remaining 10 items are now nondefective. Thus by the multiplication theorem,

$$p = \frac{8}{12} \cdot \frac{7}{11} \cdot \frac{6}{10} = \frac{14}{55} \approx 0.25$$