UNIT 9 ELEMENTARY PROBABILITY*

Structure

9.0 OBJECTIVES

After going through this unit you should be in a position to:

- explain the concept of probability,
- explain the laws of probability including Bayes' Theorem,
- solve numerical problems in probability and mathematical expectations.

9.1 INTRODUCTION

Often we make statements like

> It may rain tomorrow.
> There is a fair chance that Team A wins the match.
> It is unlikely that Mr. X becomes the President.
> Mr. Y probably met Mr. Z.

We can see that all these statements are characterised by an element of uncertainty. For example in the first statement, we are not sure that it will rain tomorrow. Similarly in the last statement, we exactly don't know whether Mr. Y met Mr. Z or not. Any statement, in which there is an element of uncertainty about the occurrence of some event, is called a *probability statement*. Thus, all the above statements are probability statements. Suppose in connection with the first probability statement, one asks, "What is the chance of rain tomorrow?"

We may have to come out with an answer like

* Dr. R A Chaudhury (Retd.), Kirorimal College, University of Delhi

'There is a 75% chance of rain tomorrow'.

Now, we are not only making a probability statement but also giving a relative measure for the degree of certainty (and implicitly that of uncertainty) associated with the event of rain tomorrow. Thus, the degree of certainty of rain is given a relative value of 75% and at the same time the uncertainty associated with rain is implicitly given a relative value of 25%. Suppose we have a scale to measure the degree of certainty. In this scale the degree of certainty will vary from 0% to 100%. This scale also implicitly measures the degree of uncertainty. Thus, if one is sure about the non-occurrence of an event, it will have 0% chance or certainty. At the same time, it will have 100% non-chance or uncertainty. Similarly, if one is sure about the occurrence of an event, it will have 100% chance or certainty and 0% non-chance or uncertainty. However, we should note here that both the statements of 100% occurrence (so, 0% non-occurrence) and 0% occurrence (so, 100% non-occurrence) of an event are statements without any element of uncertainty and hence, in a strict sense, are not probability statements.

The relative measure of the degree of certainty with which an event can occur can be termed as the *probability* of the event. Conventionally, this degree is measured relative to 1. Thus, a 30% chance of the occurrence of an event will have a probability of 0.3. Similarly, a 75% chance of rain tomorrow can be stated as

'The probability of rain tomorrow is 0.75'.

A relevant question is: How can we obtain the probability of an event? In the next section, we shall deal with this question.

9.2 DEFINITION OF PROBABILITY

It is clear that the problem of obtaining the probability is essentially a problem of measuring the degree of certainty of occurrence or non-occurrence of an event. Mathematicians have had different perceptions about the degree of certainty of an event and accordingly various definitions of probability have been given. These definitions suggest procedures for obtaining the probability of an event. In this Unit, we shall consider three such definitions. They are (i) the *classical or mathematical definition*, (ii) the *relative frequency definition* or the *statistical definition,* and (iii) the *modern definition* or the *axiomatic approach to probability*.

9.2.1 Classical Definition

The classical definition of probability is based upon certain concepts. Let us first understand these.

a) *Statistical experiment* - An experiment having more than one possible outcome is called a statistical experiment. A statistical experiment is also known as a *trial.* Thus, tossing a coin to see whether it results in a head or a tail is a trial. Certain statements implying more than one possible situation can also be termed as trials. For example, all the four statements given at the beginning of Section 9.1 are trials or statistical experiments.

b) *Event* - A possible outcome of a trial is called an event. Thus, head is an event that may result from tossing a coin. Similarly, the occurrence of five or the occurrence of an odd number is a possible event of the trial of throwing a dice. The latter example indicates that an event may consist of one or more possible outcomes of an experiment. The event of getting an odd number, in fact, consists of three possible outcomes of rolling a dice. We should note that an event consisting of only one possible outcome is often called an *elementary event*.

c) *Exhaustive events* - A set of events is said to be exhaustive, if it includes all possible outcomes of a trial. For example, the tossing of a coin can result in either a head or a tail and nothing else. Thus, the set {head, tail} is an exhaustive set of events associated with the trial of tossing a coin. Consider another example. We know, a dice has six sides and each side has dots that vary from 1 to 6. When the dice is thrown, it shows up a side with a given number of dots. If the occurrence of a side with a given number of dots is an event, the set {1, 2, 3, 4, 5, 6}, where each number represents the number of dots on a side of a dice, is an exhaustive set of events. The number of elements in an exhaustive set of events is known as the number of *cases* of the trial.

d) *Favourable events* - Such *cases* that support the occurrence of an event are said to be cases favourable to that event. Suppose a dice is thrown to see if it shows up a face with an even number of dots. In this trial, the sides with 2, 4 and 6 dots are all cases that favour the occurrence of the event of a side with an even number of dots.

e) *Equally likely events* - If in a trial, the chance of the occurrence of any possible event is the same, the events are said to be equally likely. Suppose a dice is thrown. If we feel that each of the six sides has an equal chance to show up, the possible six events are then equally likely.

f) *Mutually exclusive events* - If in a trial, the occurrence of an event rules out the simultaneous occurrence of any other possible event, the events are said to be mutually exclusive. We know, the toss of a coin results in either a head or a tail. Thus, the events of a head and a tail in the toss of a coin are mutually exclusive.

Now we are in a position to understand the classical definition of probability. The definition states:

If a trial can result in n mutually exclusive, equally likely and exhaustive outcomes and out of which m outcomes are favourable to an event A, the probability of A, denoted by P(A), is then $P(A) = \frac{m}{n}$.

It is clear that if A is an impossible event, that is, none of the n possible outcomes favours the occurrence of the event A, we have $m = 0$. The probability of A in that case is $P(A) = \frac{m}{n} = \frac{0}{n} = 0$

On the other hand, if A is a certain event, that is, all of the n possible outcomes favour the occurrence of the event A, we have $m = n$. The probability of A in that case is $P(A) = \frac{m}{n} = \frac{n}{n} = 1$

We shall now consider some applications of the classical definition for the computation of probability.

Example 9.1

What is the probability of getting a head in the toss of a fair coin?

We know that the toss of a coin can result in either a head or a tail and nothing else. Hence, these two events constitutes a set of exhaustive events. If the toss results in head, it simultaneously cannot result in tail and vice versa. Thus, the two events are mutually exclusive. Finally, if we have nothing to suspect the behaviour of the coin, both head and tail have the same chance of occurrence in the toss. So, the two events are equally likely. In this way, when all the conditions of the classical definition are satisfied, we can proceed with the solution of the problem.

The number of exhaustive outcomes $n = 2$ (head and tail)

The number of outcomes favouring the required event (head) $m = 1$

If P (H) is the probability of the occurrence of head, then $P(H) = \frac{m}{n} = \frac{1}{2}$

From now onwards, we shall proceed straight with the solution. However, you should satisfy yourselves that all the conditions of the classical definition are satisfied.

Example 9.2

A fair dice is thrown. What is the probability that either 1 or 6 will show up?

A dice has six faces with 1, 2, 3, 4, 5 and 6 dots printed on them and any one of these faces will show up when the dice is thrown. Thus the number of exhaustive outcomes $n = 6$. Now, the face with 1 dot favours the required event and the face with 6 dots also satisfies the required event. So, the number of outcomes favouring the required event is $m = 2$. If $P(1 \; or \; 6)$ is the probability of either 1 or 6 then

$$P(1 \; or \; 6) = \frac{m}{n} = \frac{2}{6} = \frac{1}{3}$$

Example 9.3

An unbiased coin is tossed twice. What is the probability of getting a head at least once?

If H stands for head and T stands for tail, there are four possible outcomes. They are

(H, H) (H, T) (T, H) (T, T)

Thus, $n = 4$ here and the number of outcomes favourable to the required event of at least one head, $m = 3$. Hence

$$P\ (at\ least\ one\ head) = \frac{m}{n} = \frac{3}{4}$$

Limitations of the Classical Definition

The classical definition has some serious drawbacks. They are

a) The classical definition can be applied only if various outcomes of the trials are equally likely or equally probable. But in practice the outcomes need not be always equally likely. For example, if a coin is biased in favour of head, the classical definition fails to give the probability of a head or a tail.

b) The classical definition is valid for a finite number of outcomes of a trial. It fails when the number of outcomes becomes infinity. In fact, even in the case of a finite number of outcomes, it may not be practically feasible to enumerate all the cases.

c) The classical definition is 'circular' in the sense that while defining probability, the definition uses the term 'equally likely' which pre-supposes the knowledge of the concept of probability.

Check Your Progress 1

1) A box contains 4 white balls and 6 red balls. A ball is drawn without looking into the box. What is the probability that it is a white ball?

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

2) A six-faced dice is thrown. What is the probability of getting an even number?

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

3) A coin is tossed twice. What is the probability of getting either two heads or two tails?

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

4) A card is drawn from a pack of 52 cards. What is the probability of not getting a king?

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

9.2.2 Relative Frequency or Statistical Definition

Another definition that has often been used is the relative frequency definition of probability. If we repeat a trial and observe the occurrence of an event, we shall see that as the number of trials is progressively increased, the ratio of the number of times a particular event occurs to the total number of trials tends to stabilise at a particular value. Now, the number of times an event occurs is its frequency and when this frequency is divided by the total number of trials, we get the relative frequency of the event. Thus in other words, when the number of trials becomes sufficiently large, the relative frequency of an event tends to a limit. According to the relative frequency definition, this limiting value is the probability of the event under consideration. Suppose, we repeat the experiment of tossing a coin and observe the number of times head occurs. We shall find that as we increase the number of tosses from say, 10 to 100 to 1000 to 10000 and so on, the relative frequency of head will gradually stabilise at $\frac{1}{2}$. Thus, the probability of head in the toss of a fair coin is $\frac{1}{2}$.

Mathematically, if n is the total number of trials out of which, an event A occurs m times, the probability of A

$$P(A) = \lim_{n \to \infty} \frac{m}{n}$$

9.2.3 Axiomatic Approach to Probability

The main limitation of the classical as well as the relative frequency definitions is that these definitions preclude a rigorous mathematical treatment of the subject of probability. This limitation has been taken care in the modern definition.

Before presenting the modern or axiomatic definition, it is necessary to grasp the following concepts.

a) *Sample Space* - It is the set of all possible (or exhaustive) outcomes of a trial. The sample space of a trial can be denoted by S and is given by $S = \{e_1, e_2, \ldots, e_n\}$, where, $e_1, e_2, \ldots, e_n$ are n elementary events.

   If the trial consists of tossing a coin, then the sample space will be $S = \{H, T\}$. Similarly, when a dice is rolled, the sample space is given by $S = \{1, 2, 3, 4, 5, 6\}$.

   The elements of a sample space can also be *ordered pairs*. For example, the sample space of the simultaneous toss of two coins is $S = \{ (H, H), (H, T), (T, H), (T, T) \}$. Further, a sample space can be *finite* or *infinite* depending upon whether it consists of finite or infinite number of elements.

b) *Event* - An event is any *subset* of sample space. For example, if A denotes an event that an odd number appears on a dice, then $A = \{1, 3, 5\}$. Again, the event of the occurrence of at least one head when two coins are tossed simultaneously is given by, say, $B = \{ (H, H), (H, T), (T, H) \}$.

c) *Occurrence of an Event* - An event is said to have occurred whenever the outcome of a trial belongs to the relevant event-subset. Thus, when we roll a dice and get 1, we say that event A has occurred. Based upon this, we can say that the sample space of a trial is certain to occur.

According to the modern definition, the probability of an event A, denoted by P(A) is a *real valued set function* that associates a real value P(A) corresponding to any *subset A* of the sample space S.

In order that P(A) is the probability of an event A, it must satisfy the following restrictions. These restrictions are also known as the *axioms of probability theory*.

1. The probability of an event A, in a sample space S, is a non-negative real number less than or equal to unity, i.e., $0 \le P(A) \le 1$.

2. The probability of an event, that is certain to occur, is unity. Since S is certain to occur, this implies that $P(S) = 1$.

3. If A_1, A_2 and A_3 are mutually exclusive events in a sample space S, then

$$P\left(A_1 \cup A_2 \cup A_3\right) = P\left(A_1\right) + P\left(A_2\right) + P\left(A_3\right).$$

   The above relation can be generalised to any number of events.

We should note that in the sample space $S = \{e_1, e_2, \ldots, e_n\}$, the elementary events, $e_1, e_2, \ldots, e_n$ are mutually exclusive. Thus on the basis of the third axiom, we can say that

$$P(S) = \sum_{i=1}^{n} P\left(e_i\right)$$

Verbally, the probability of sample space is equal to the sum of the probabilities of its elementary events. In a similar way, we can state that the probability of an event is equal to the sum of its elementary events.

Thus, to find the probability of an event, we must know the probability of the occurrence of its elementary events. This can be done in any of the following three ways.

1. In the absence of any information regarding the occurrence of various elementary events, it is reasonable assume them to be equally likely. Thus, we can assign equal probability to each of the elementary events.

   Since $P(S) = \sum_{i=1}^{n} P(e_i) = 1$, therefore

   $$P(e_i) = \frac{1}{n} \quad (i = 1, 2, \ldots, n)$$

   If there are m elementary events in the event A, then

   $$P(A) = \left(\frac{1}{n} + \frac{1}{n} + \cdots + \frac{1}{n}\right) m \text{ times}$$

   $$= \frac{m}{n} = \frac{\text{number of elements in } A}{\text{number of elements in } S}$$

   This is nothing but the classical approach to probability.

2. Another way of assigning probability to various elementary events is to perform an experiment a large number of times. The relative frequencies of various elementary events can be taken as their respective probabilities if n is sufficiently large. This method uses the statistical definition of probability discussed before.

3. The probabilities of various elementary events can also be assigned by the person performing the experiment, on the basis of her experience and expectations. For example, one may ask you to specify the probability of rain today. You may be tempted to specify a higher value, say 0.8, if the day falls in the rainy season and so on. This approach to probability is particularly useful to managers engaged in taking various business decisions.

In practice we encounter many situations which involve a combination of events. In such situations we need to combine the probabilities of these events. In this context, we discuss two important laws of probability below.

9.3 PROBABILITY LAWS

Before considering various probability laws, let us be familiar with certain notations.

a) If A and B are two events, then $P(A \cup B)$ or $P(A + B)$ denotes the probability that either A occurs or B occurs or both occur simultaneously. It can also be interpreted as the probability of the occurrence of at least one of the two events *A* and *B*.

b) $P(A \cap B) or\, P(AB)$ denotes the probability of the simultaneous occurrence of both *A* and *B*.

c) $P(A/B)$ denotes the *conditional probability* of the occurrence of *A* given that *B* has already occurred.

9.3.1 Addition Law

This law states that the probability of the occurrence of at least one of the two events (i.e., either *A* or *B* or both) is equal to the probability of *A* plus the probability of *B* minus the probability of both *A* and *B*.

Using notations, we can say

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \qquad \ldots(9.1)$$

We discuss below the modification to (9.1) in certain special cases.

a) *Mutually Exclusive Events*: Now suppose, *A* and *B* are mutually exclusive, that is, the occurrence of *A* precludes the occurrence of *B* and vice versa; then, the two events cannot occur simultaneously and $P(A \cap B) = 0$. Thus, for two mutually exclusive events *A* and *B*, probability of the occurrence of either *A* or *B* is equal to the probability of A plus the probability of *B*.

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

b) *Exhaustive Events*: Again, if *A* and *B* are the only possible outcomes of a trial (i.e., *A* and *B* are exhaustive events), then the occurrence of either *A* or *B* is a certainty. We know that the probability of a event that is certain to occur is 1. Hence, in that case

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 1$$

or $P(A \cup B) = P(A) + P(B) = 1$ (when *A* and *B* are mutually exclusive)

c) *Complementary events:* Suppose *A* is a possible outcome of some trial. It is then clear that the trial either results in the occurrence of *A* or the non-occurrence of *A*. Thus, *A* and 'not *A*' exhaust all the possible outcomes of any trial. So, if $\overline{A}$ denotes 'not *A*', we have

$$P(A) + P(\overline{A}) = 1$$

or $P(\overline{A}) = 1 - P(A)$

Here, $\overline{A}$ is called *complement* to the event *A*. Thus, the sum of probabilities of any event and its complement is always equal to 1.

9.3.2 Multiplication Law

This law states that the probability of the simultaneous occurrence of the two events A and B is equal to the product of

(i) the probability of A and the conditional probability of B given that A has already occurred

or

(ii) the probability of B and the conditional probability of A given that B has already occurred.

In symbols,

$$P(A \cap B) = P(A) \cdot P(B/A) = P(B) \cdot P(A/B) \qquad \ldots(9.2)$$

Using the multiplication law, we can find the *conditional probabilities*

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

and

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

In case of multiplication law, certain modification is required in case of *independent events*.

Suppose that the occurrence of B does not depend upon the occurrence of A and vice versa, then the two events A and B are said to be mutually independent. In this case the two conditional probabilities $P(B/A)$ and $P(A/B)$ are equal to their respective non-conditional simple probabilities. Hence, for independence

$$P(B/A) = P(B) \quad \text{and}$$
$$P(A/B) = P(A)$$

Thus, for two independent events A and B, the probability of their simultaneous occurrence is the product of their respective probabilities.

$$P(AB) = P(A) \cdot P(B) = P(B) \cdot P(A)$$

Let us now consider some examples on the application of the probability laws.

9.3.3 Applications of Probability Laws

Let us work out some problems so that you get a fair idea of the application of the above mentioned probability laws.

Example 9.4

A dice is thrown. What is the probability of getting either 1 or 6?

It is clear that in a single throw the two events of 1 and 6 cannot occur together. Hence, the two events are mutually exclusive. Thus,

$$P(\textit{either 1 or 6}) = P(1) + P(6) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}.$$

Example 9.5

One card is drawn from a pack of 52 cards. The card is *not replaced* in the pack and another card is drawn. What is the probability that both the cards are spade?

Here the first event is drawing a spade and the second event is drawing another spade given that the first card is a spade. Thus the second event is a conditional event. Let $P(A)$ be the probability of the first event and $P(B/A)$ be the probability of the second event given the occurrence of the first event. Now, $P(A)=\frac{13}{52}=\frac{1}{4}$. When the first card is a spade and is not replaced, there are 12 spades left in a pack of 51 cards. So, $P(B/A)=\frac{12}{51}$. Hence, the required probability is

$$P(A\cap B)=P(A)\cdot P(B/A)=\frac{13}{52}\cdot\frac{12}{51}=\frac{1}{4}\cdot\frac{12}{51}=\frac{3}{51}$$

We should note that when the card is not replaced after the first drawing, the two events are not independent as the probability of the occurrence of the second event depends upon the probability of the occurrence of the first event.

Example 9.6

One card is drawn from a pack of 52 cards. The card is *replaced* in the pack and another card is drawn. What is the probability that both the cards are spade?

In this example, the card is replaced after the first drawing. Thus, when the second card is drawn, there are 13 spades in a pack of 52 cards. As a result, the probability of the second card being a spade does not depend upon whether the first card drawn is a spade or not. Hence, the two events are independent here. If $P(B)$ is the probability of the second card being a spade, in this case

$$P(B/A)=P(B)=\frac{13}{52}=\frac{1}{4}.$$

Thus, the required probability is

$$P(A\cap B)=P(A)\cdot P(B/A)=P(A)\cdot P(B)=\frac{13}{52}\cdot\frac{13}{52}=\frac{1}{4}\cdot\frac{1}{4}=\frac{1}{16}$$

Example 9.7

A dice is thrown. What is the probability of getting a number less than 5 or an odd number?

Let A be the event of a number less than 5 and B be the event of an odd number. We should note here that the two events are not mutually exclusive as a number can be both less than 5 and an odd number. So the required probability is obtained by applying the formula

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

Now in a dice, out of 6 numbers, there are 4 numbers (1, 2, 3, and 4) less than 5. Therefore,

$$P(A) = \frac{4}{6} = \frac{2}{3}.$$

Again, there are 3 odd numbers (1,3 and 5) out of a possible six numbers. So

$$P(B) = \frac{3}{6} = \frac{1}{2}$$

Suppose $P(B/A)$ is the probability of an odd number given it is less than five. Then

$$P(B/A) = \frac{2}{4} = \frac{1}{2}$$

Now

$$P(AB) = P(A)\cdot P(B/A) = \frac{2}{3}\cdot\frac{1}{2} = \frac{1}{3}$$

Thus, the probability of getting a number less than 5 or an odd number is

$$\frac{2}{3} + \frac{1}{2} - \frac{1}{3} = \frac{5}{6}$$

Example 9.8

If A and B are two events such that $P(A) = \frac{2}{3}$, $P(\overline{A} \cap B) = \frac{1}{6}$ and $P(A \cap B) = \frac{1}{3}$. Find $P(B)$, $P(A \cup B)$, $P(A/B)$, $P(B/A)$, $P(\overline{A} \cup B)$ and $P(\overline{A} \cap \overline{B})$. Also examine whether A and B are

a) Equally likely

b) Exhaustive

c) Mutually exclusive

d) Independent.

We can write

$$P(B) = P(\overline{A} \cap B) + P(A \cap B) = \frac{1}{6} + \frac{1}{3} = \frac{3}{6} = \frac{1}{2}$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{2}{3} + \frac{1}{2} - \frac{1}{3} = \frac{5}{6}$$

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{1}{3}\cdot 2 = \frac{2}{3}$$

$$P(B/A) = \frac{P(A \cap B)}{P(A)} = \frac{1}{3}\cdot\frac{3}{2} = \frac{1}{2}$$

$$P(\overline{A} \cup B) = P(\overline{A}) + P(B) - P(\overline{A} \cap B) = \frac{1}{3} + \frac{1}{2} - \frac{1}{6} = \frac{2}{3} \qquad \left[\because P(\overline{A}) = 1 - P(A)\right]$$

$$P(\overline{A} \cap \overline{B}) = 1 - P(A \cup B) = 1 - \frac{5}{6} = \frac{1}{6} \qquad \text{(Using the concept of complementary event)}$$

a) Since $P(A) \neq P(B)$, A and B are not equally likely.

b) Since $P(A \cup B) \neq 1$, A and B are not exhaustive.

c) Since $P(A \cap B) \neq 0$, A and B are not mutually exclusive.

d) Since $P(A) \cdot P(B) = P(A \cap B)$, A and B are independent.

Check Your Progress 2

1) A student takes Mathematics and English tests. His independent chances of passing the two tests are $\frac{2}{3}$ and $\frac{3}{4}$ respectively. What is the probability that

   a) he passes at least one test?

   b) he fails in both the tests?

……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………

2) Two cards are drawn from a pack of 52 cards. What is the probability

   a) that both the cards are kings when the first card is *replaced* before the second card is drawn?

   b) that both the cards are spades when the first card is *not replaced* before the second card is drawn?

……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………

3) There are two urns. The first urn contains 7 white balls and 3 red balls. The second urn contains 4 white balls and 6 red balls. An urn is selected at random and a ball is drawn. What is the probability that the first urn is selected and a red ball is drawn from it?

……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………

4) The probabilities that A and B speak the truth independently are $\frac{1}{2}$ and $\frac{1}{3}$ respectively. If they make the same statement, what is the probability that the statement made by them is a true one?

………………………………………………………………………………………

………………………………………………………………………………………

………………………………………………………………………………………

………………………………………………………………………………………

………………………………………………………………………………………

9.4 BAYES' THEOREM

Let A_1, A_2, and A_3 be three mutually exclusive and exhaustive events and there be an event D which can occur in conjunction with any of them. If D actually happens, then the conditional probability of the occurrence of A_i (i = 1, 2, 3) given D, is given by

$$P\left(A_i / D\right) = \frac{P\left(A_i \cap D\right)}{P(D)} = \frac{P\left(A_i\right) \cdot P\left(D / A_i\right)}{P(D)}$$

where

$$P(D) = \sum_{i=1}^{3} P\left(A_i \cap D\right) = \sum_{i=1}^{3} P\left(A_i\right) \cdot P\left(D / A_i\right)$$

We should note that the above result can be generalised to any number of mutually exclusive and exhaustive events.

Let us consider some practical applications of Bayes' Theorem.

Example 9.9

In a factory that produces bolts there are three machines *A, B* and *C*. They manufacture 25%, 35% and 40% of total output respectively. However, 5%, 4% and 2% of their respective output are defective. A bolt is drawn at random from a day's output and is found to be defective. What is the probability that it was produced by (i) machine *A,* (ii) machine *B,* and (iii) machine *C* ?

Since a day's output consists of the bolts produced by all the three machines, the probability that a bolt selected at random is produced by machine A is given by $P(A) = 0.25$. Similarly we have, $P(B) = 0.35$ and $P(C) = 0.40$. Further, let D be the event that the bolt is defective. Since machine A produces 5% defective bolts, we have $P(D / A) = 0.05$. Similarly, $P(D / B) = 0.04$ and $P(D / C) = 0.02$.

Thus

$$P(D) = P(A) \cdot P(D / A) + P(B) \cdot P(D / B) + P(C) \cdot P(D / C)$$

$$= 0.25 \times 0.05 + 0.35 \times 0.04 + 0.40 \times 0.02 = 0.0345$$

The probability that the bolt is manufactured by machine A given that it is defective

$$P(A/D) = \frac{P(A) \cdot P(D/A)}{P(D)} = \frac{0.25 \times 0.05}{0.0345} = \frac{0.0125}{0.0345} = 0.362$$

Similarly,

$$P(B/D) = \frac{P(B) \cdot P(D/B)}{P(D)} = \frac{0.35 \times 0.04}{0.0345} = \frac{0.0140}{0.0345} = 0.406$$

and

$$P(C/D) = \frac{P(C) \cdot P(D/C)}{P(D)} = \frac{0.40 \times 0.02}{0.0345} = \frac{0.0080}{0.0345} = 0.232$$

There is an alternative method you can pursue. You can determine the above probabilities by making the following table. The three events A, B and C have been renamed as A_1, A_2, and A_3 respectively.

A_i	A_1	A_2	A_3	Total
$P(A_i)$	0.25	0.35	0.45	1.00
$P(D/A_i)$	0.05	0.04	0.02	
$P(D \cap A_i)$	0.0125	0.014	0.008	
$P(A_i/D) = \frac{P(D \cap A_i)}{P(D)}$	0.362	0.406	0.232	1.00

Note that the probabilities $P(A_1)$, $P(A_2)$ and $P(A_3)$, which are known before conducting the trial, are known as *priori* probabilities. The conditional probabilities of A_1, A_2, and A_3 i.e., $P(A_1/D)$, $P(A_2/D)$ and $P(A_3/D)$, after the result of the trial is known; are termed as *posterior* probabilities.

To begin with the analysis of a problem, the manager of a firm assigns probabilities to certain events on subjective basis, i.e., based on his experience and expectation. These probabilities are the priori probabilities. Then the trial is conducted. Subsequent to the trial, the priori probabilities are revised on the basis of the occurrence of certain event like D to obtain posterior probabilities. Again in the next round, these posterior probabilities can be taken as priori probabilities and the whole procedure may be repeated to revise the posterior probabilities. Such a revision can be repeated a number of times. Generally, after a certain number of revisions the posterior probabilities tend to stabilise and the subjective probabilities tend to become objective probabilities. Thus, Bayes' Theorem proves to be very useful for the analysis of business phenomena.

Example 9.10

The probability that the product of a company will be successful given that the result of the survey is favourable is 0.6 and the probability of its being successful with unfavourable survey is 0.3. If the probability that the survey shows a favourable result is 0.7, find the probability that (i) the product is successful; (ii) the result of the survey is favourable given that the product is successful, and (iii) the result of the survey is unfavourable given that the product is successful.

Let S denote the event that the product of the company is successful and F denote the event that the survey result is favourable. Let $\overline{S}$ and $\overline{F}$ be the events denoting the negation of the respective events.

In terms of notations we are given

$$P(S/F) = 0.6,\ P(S/\overline{F}) = 0.3 \text{ and } P(F) = 0.7$$

Thus we have

$$P(\overline{F}) = 1 - 0.7 = 0.3$$

(i) The probability that the product is successful is given by

$$P(S) = P(S \cap F) + P(S \cap \overline{F})$$

$$= P(F) \cdot P(S/F) + P(\overline{F}) \cdot P(S/\overline{F})$$

$$= 0.7 \times 0.6 + 0.3 \times 0.3 = 0.51$$

(ii) $P(F/S) = \dfrac{P(F \cap S)}{P(S)} = \dfrac{0.42}{0.51} = 0.824$

(iii) $P(\overline{F}/S) = \dfrac{P(\overline{F} \cap S)}{P(S)} = \dfrac{0.09}{0.51} = 0.176$

Note that $P(\overline{F}/S) = 1 - P(F/S)$

Check Your Progress 3

1) A talcum powder manufacturing company had launched a new type of advertisement. The company estimated that a person who comes across their advertisement will buy their product with a probability of 0.7 and those who do not see the advertisement will buy the product with a probability of 0.3. If in an area of 1000 people, 70% had come across the advertisement, find the probability that a person who buys the product

   (a) has not come across the advertisement and.

   (b) has come across the advertisement?

   …………………………………………………………………………………

   …………………………………………………………………………………

   …………………………………………………………………………………

   …………………………………………………………………………………

2) An insurance company insured 2000 scooter drivers, 4000 car drivers and 6000 truck drivers. The probability of an accident is 0.01, 0.03 and 0.15 in the respective categories. One of the insured drivers meets with an accident. What is the probability that the person is a scooter driver?

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

……………………………………………………………………………...................

……………………………………………………………………………...................

9.5 LET US SUM UP

In ordinary parlance, probability refers to chance. In statistics, however, we go deeper than this. Here, we not only consider the uncertainty involved in the occurrence of an event but also try to quantify it. Thus, probability is a quantitative measure of chance. In this Unit, we have considered three different approaches to probability, namely, the classical approach , the relative frequency approach and the axiomatic approach. The probabilities of compound events are essentially governed by two laws. They are the addition law and the multiplication law. Finally, Bayes' Theorem provides a framework for revising the probabilities on the basis of the occurrence and non-occurrence of certain events. This revision is very useful for business decisions.

9.8 ANSWERS OR HINTS TO CHECK YOUR PROGRESS EXERCISES

Check Your Progress 1

1) $\frac{2}{5}$

2) $\frac{1}{2}$

3) $\frac{1}{2}$

4) $\frac{12}{13}$

Check Your Progress 2

1) a) $\frac{11}{12}$, b) $\frac{1}{12}$

2) a) $\frac{1}{169}$, b) $\frac{1}{17}$

3) $\frac{3}{20}$

4) $\frac{1}{3}$

Check Your Progress 3

1) $\frac{9}{58}, \frac{49}{58}$

2) $\frac{1}{52}$

ignou
THE PEOPLE'S
UNIVERSITY

UNIT 10 DISCRETE PROBABILITY DISTRIBUTIONS*

Structure

10.0 OBJECTIVES

After going through this unit you should be able to:

- explain the meaning of a random variable
- explain the concept of a probability distribution,
- distinguish between a discrete probability distribution and a continuous probability distribution, and
- explain the binomial and the Poisson distribution.

10.1 INTRODUCTION

In Unit 9, we discussed about probability of the occurrence of an event. In that unit an event was defined as the set of one or more possible outcomes of a chance experiment. The outcomes of such a chance experiment can be related to the concept of a random variable. In the present unit, we shall consider probability in the context of a random variable and understand the notion of a probability distribution. Any probability distribution is based upon the behaviour of some random variable. In this unit, we shall define a random variable and distinguish between a discrete random variable and a continuous random variable. Then, in the context of discrete random variables, we shall discuss two important discrete probability distributions. They are the binomial distribution and the Poisson distribution.

* Dr. Anup Chatterjee (retd.), ARSD College, University of Delhi

10.2 RANDOM VARIABLE

Before presenting the formal definition of a random variable, let us intuitively try to understand the concept of a random variable. As mentioned in the introduction, a random variable is related to the outcomes of a chance experiment. Such a chance experiment is also known as a random experiment. Let us consider an example.

Suppose a coin is tossed. There are two possible outcomes; a head (H) and a tail (T). In the earlier unit, we have discussed the concept of a sample space. The sample space of this experiment consists of the outcomes head and tail. If S denotes the sample space, then

$$S = (H, T)$$

In this experiment, we are not sure whether a head will result or a tail. This is an example of a chance experiment or a random experiment. Now suppose, we assign a number 0 to the occurrence of tail (T = 0) and a number 1 to the occurrence head (H = 1). Let us define a variable X that refers to the occurrence of an outcome. Then the variable and its possible values can be written as

$$X = (0, 1)$$

However, there is an important difference between this variable and our common notion of a variable. Here, the value that the variable takes depends upon the outcome of the chance or random experiment that we are considering. In other words, we are not sure, whether as a result of the experiment; the variable will take the value 0 or 1. We can only attach some probabilities to these values. These probabilities depend upon the chances of the occurrence of the different outcomes of the experiment. If in our example, for instance, the coin is unbiased, the probability of the occurrence of tail is $\frac{1}{2}$ and that of head is also $\frac{1}{2}$; because, both the outcomes have an equal chance to occur. Accordingly, we attach a probability of $\frac{1}{2}$ to both 0 and 1. In the case of the conventional notion of a variable, on the other hand, no such probability is attached to a value taken by the variable.

From the above discussion we can say that a random variable is *a variable that takes different values with some probabilities*. Thus, the variable X referring to the possible outcomes of tossing a coin, is an example of a random variable.

We will use the following notations: Let X be a random variable and it assumes values $x_1, x_2, x_3, ...x_n$. The corresponding probabilities are $p_1, p_2, p_3, ...p_n$. Thus, $p(X = x_1) = p_1$.

Example 10.1

Let us consider an experiment of simultaneous tossing of two coins. The sample space of the experiment is

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

If we define a random variable X as the number of heads obtained, then $X = 2$ corresponds to the outcome (H, H); $X = 1$ corresponds to the outcomes (H, T) and (T, H); and finally, $X = 0$ corresponds to the outcome (T, T). Thus X can take three possible values i.e., 0, 1 and 2.

$X = (0, 1, 2)$

Example 10.2

As another example, we consider the roll of a dice. The sample space is $S = \{1, 2, 3, 4, 5, 6\}$. A random variable X can be defined such that it takes a value equal to 0 when an odd number appears on the dice and 1 when an even number appears. Thus

$X = (0, 1)$

In the tossing of two coins experiment, we can also define a random variable in monetary terms. For instance, we may decide to pay a player Rs. 10/- if two heads are obtained, Rs. 5/- if one head is obtained and ask the player to pay Rs. 8/- (i.e., we pay Rs. –8/-) if no head is obtained. Here X is a random variable denoting the amount of payment that can be made to a player. Thus

$X = (10, 5, -8)$

A random variable can be either discrete or continuous.

(i) *Discrete Random Variable*: When the sample space of an experiment is discrete, the corresponding random variable will also be discrete, i.e., it will take certain isolated values. The random variables discussed above are examples of discrete random variable.

(ii) *Continuous Random Variable*: We know that a continuous variable can take any value in an interval. Accordingly, a continuous random variable is defined when the accompanying sample space is also continuous. In the next unit, we shall discuss the concept of the normal variable which is an example of a continuous random variable.

10.3 PROBABILITY DISTRIBUTION

Let us begin with a definition of probability distribution. It is defined as a statement about the possible values of a random variable along with their respective probabilities.

Let us take a concrete example of probability distribution. In the earlier example of tossing two coins, we defined a random variable X as the number of heads. Further, X took three values, viz., 0, 1 and 2. Assuming that the two coins are unbiased, we can write

$p\ (X=0) = \frac{1}{4}$,

$p(X=1)=\frac{1}{2}$ [i.e., the probability of the occurrence of (H, T) or (T, H)]

$p\ (X=2)=\frac{1}{4}$.

These probabilities along with the corresponding values of the random variable written in a tabular form constitute *the probability distribution of the random variable X* where X is the number of heads. It is shown in Table 10.1.

Table 10.1: Probability Distribution of the Number of Heads Obtained in Tossing of Two Unbiased Coins

Number of Heads (x)	Probability $p(x)$
0	$\frac{1}{4}$
1	$\frac{1}{2}$
2	$\frac{1}{4}$

In the above example, the events of getting no head (X = 0), one head (X = 1),and two heads (X = 2) *exhaust* all the possibilities (this means, there is no other possible outcome than the above three). Thus, the probability distribution resulting from the above experiment has enumerated all the possible values of the random variable X and assigned specific probabilities to them. We can see that the sum of these probabilities equals 1.

Probability distribution can be of two types: Discrete Probability Distribution and Continuous Probability Distribution.

10.3.1 Discrete Probability Distribution

We have already seen that the probability distribution for a random variable describes how the probabilities are distributed over the values of the random variable. Now, for a discrete random variable, the probability distribution is defined by a function called *probability mass function*, denoted by $p(x)$. This probability mass function provides the probability for each value of the discrete random variable. In fact, the probability distribution of the number of heads in tossing two coins that we have presented in Table 10.1 is an example of a discrete probability distribution.

We can consider another example of a discrete probability distribution. Suppose we observe the number of children per household in a locality. Here, we can consider the number of children as a discrete random variable. A discrete probability distribution for the number of children per household can be constructed by computing the relative frequencies for the possible values of this random variable. Such a probability distribution is shown in Table 10.2.

Table 10.2: Probability Distribution of the Number of Children per Household

Number of Children (x)	$p(x)$
0	0.10
1	0.15
2	0.23
3	0.25
4	0.14
5	0.13

Thus, the set of ordered pairs [*x, p(x)*] is called the probability distribution of a discrete random variable *X* or the discrete probability distribution.

Since the values *p(x)* are all probabilities and a value *x* of the random variable will always occur, the probability mass function should satisfy the following two conditions

1) Probability of an event cannot be negative, i.e., for any value of *X*

   $p(x) \geq 0$

2) Probabilities of all possible outcomes sum to unity, i.e.,

   $\sum_{all\,x} p(x) = 1$

Let us work out some problems on discrete probability distribution.

Example 10.3

Is the following a valid probability mass function? $p(x) = \frac{x^3}{2}, x = -1, 0, 1$

Let us find out the probability of *x*, when *x* assumes the specified values (−1, 0 and 1).

When $x = -1$

$$p(x) = p(-1) = -\frac{1^3}{2} = -\frac{1}{2} < 0 .$$

But we know that probability of an event cannot be negative. Thus, the first condition of a probability mass function is violated. Hence, the given function is not a valid probability mass function.

Example 10.4

Given a function $p(x)=\frac{k}{x}$, x = 3, 4, 5 and k is a constant. Find k such that the given function is a valid probability mass function.

From the given function, we have

$$p(3)=\frac{k}{3}$$

$$p(4)=\frac{k}{4}$$

$$p(5)=\frac{k}{5}$$

For the satisfaction of the second condition of a probability mass function, we have

$$\sum p(x)=\frac{k}{3}+\frac{k}{4}+\frac{k}{5}=1$$

$$or\ k\left(\frac{1}{3}+\frac{1}{4}+\frac{1}{5}\right)=1$$

$$or\ k\cdot\frac{47}{60}=1$$

$$or\ k=\frac{60}{47}$$

When $k=\frac{60}{47}$

$$p(3)=\frac{1}{3}\cdot\frac{60}{47}=\frac{60}{141}\geq 0$$

$$p(4)=\frac{1}{4}\cdot\frac{60}{47}=\frac{60}{188}\geq 0$$

$$p(5)=\frac{1}{5}\cdot\frac{60}{47}=\frac{60}{235}\geq 0$$

Thus, for $k=\frac{60}{47}$, the first condition for a probability mass function is also satisfied.

14.3.2 Continuous Probability Distribution

A continuous random variable X has a zero probability of assuming exactly any of its values. Apparently, this seems to be a surprising statement. Let us try to understand this by considering a random variable say, weight.

Obviously weight is a continuous random variable since it can vary continuously. Suppose, we do not know the weight of a person exactly but have a rough idea that her weight falls between 60 kg and 61 kg. Now, there are an infinite number of possible weights between these two limits. As a result, by its definition, the probability of the person's assuming a particular weight say, 60.3 kg will be negligibly small almost equal to zero. But we can definitely attach some probability to the person's weight being between 60 kg and 61 kg. Thus, for a continuous random variable X, one assigns a probability to an interval and not to a particular value. Here, we look for a function $p(x)$, called the *probability density function*, such that with the help of this function we can compute the probability

$P(a < x < b)$, a and b are the limits of an interval (a, b) where, $a < b$

A probability density function is defined in such a manner that the area under its curve bounded by x-axis is equal to one when computed over the domain of X for which $p(x)$ is defined. The probability density function for a continuous random variable X defined over the entire set of real numbers R should satisfy the following conditions.

1) $p(x) \geq 0$ for all $x \in R$
2) $\int_{-\infty}^{+\infty} p(x)dx = 1$
3) $p(a < X < b) = \int_{a}^{b} p(x)dx$

Although the probability distribution of a continuous random variable cannot be presented in the form of a table like that of a discrete random variable, it can nevertheless be expressed by a specific form of the probability density function $p(x)$. We shall study some of these forms in the next unit on the theoretical distributions for continuous random variables.

10.3.3 Theoretical Distributions

We should note that a probability distribution is based upon the empirical observations associated with a probability experiment. For obtaining the relevant probability distribution, the experiment has to be repeated a very large number of times under an identical condition which may sometimes prove to be an extremely difficult task. Alternatively, by means of a formula, we can specify a probability mass function or a probability density function, as the case may be, theoretically by satisfying the conditions of the experiment. Such kind of a probability distribution is known as a *theoretical distribution*. An advantage with the theoretical distribution is that a few theoretical distributions describe many real life random phenomena. Consequently, with the help of a handful of important theoretical distributions we may get an insight into several such real life random phenomena without actually experimenting with them. A theoretical distribution can either be a discrete or a continuous one.

Later in this unit, we discuss two important *discrete theoretical distributions* that have often been employed in the statistical analysis. In the next unit, we shall study some continuous theoretical distributions.

10.4 MEAN AND VARIANCE OF A RANDOM VARIABLE

The mean of a random variable, also known as its *mathematical expectation* or *expected value* is defined as the sum of the products of the values of the random variable and the corresponding probabilities.

Thus, if X is a discrete random variable that can assume the values $x_1, x_2, x_3, \ldots,$ x_n with specific probabilities p_1 , p_2 , p_3 , …, p_n respectively, then the mathematical expectation of x is

$$E(X) = x_1p_1 + x_2p_2 + x_3p_3 + \ldots\ldots + x_np_n = \sum_{i=1}^{n} x_ip_i$$

It is interesting to note that the mathematical expectation of a random variable corresponds to the arithmetic mean of an ordinary variable. This can be easily shown for a discrete random variable. We have seen from the *relative frequency definition of probability* that the probability of an event can be interpreted as the limit of relative frequency of the occurrence of that event when the number of trials tends to infinity, i.e.,

$$p_i = \frac{f_i}{N}$$

where f_i is frequency of x_i and $N = \sum_{t-1}^{n} f_i$ is the total frequency.

Thus $E\left(X\right) = \sum_{i=1}^{N} p_ix_i = \sum_{t-1}^{n} \frac{f_i}{N} x_i$

$= \frac{1}{N}\sum_{i=1}^{N} f_ix_i = \overline{X}$, the arithmetic mean of X.

Example 10.5

A fair coin is tossed. If it is 'head', you win Rs. 20/-. If it is 'tail', you lose Rs. 10/-. What is the amount that you are expected to win or lose per toss?

Since the coin is given to be unbiased, the probability of getting a 'head' or a 'tail' is $\frac{1}{2}$. Let X be a random variable which takes the values equal to the amounts of gain and loss. So, X=20 with probability $\frac{1}{2}$ and $x = -10$ (loss can be considered to be negative gain) again with probability $\frac{1}{2}$.

Therefore, the expected amount to win or lose per toss is

$$Rs.\left[20\cdot\frac{1}{2} + \left(-10\right)\cdot\frac{1}{2}\right] = Rs.\ 5$$

A game with positive expected gain is said to be biased in favour of the player. If the expected gain is zero, the game is said to be fair. The above game can be made fair if we charge Rs. 5/- (equal to the expected value) as entry fee. The possible values of the random variable X now become 15 and −15 and the expected value $E(X) = 0$.

For a continuous random variable, the mathematical expectation takes the form of a definite integral. Thus,

$$E(X) = \int_a^b xp(x)dx$$

where, X is a continuous random variable with domain from a to b and $p(x)$ is its probability density.

10.4.1 Theorems on Mathematical Expectation

(i) The mathematical expectation of a constant is the constant itself. If c is a constant, then

$$E(c) = c$$

(ii) The mathematical expectation of the product of a constant and a random variable is the product of the constant and the mathematical expectation of the random variable. If c is a constant and X is a random variable, then

$$E(cX) = cE(X)$$

(iii) The mathematical expectation of any function of a random variable is the sum of the products of the values of the function and the corresponding probabilities of the values of the random variable. Thus if $f(X)$ is a function of a random variable X that takes the values $x_1, x_2, x_3, \ldots, x_n$ with specific probabilities $p_1, p_2, p_3, \ldots, p_n$ the mathematical expectation of $f(X)$ is

$$E[f(X)] = \sum_{i=1}^{n} f(x_i) p_i$$

We may note here that the above summation, strictly speaking, applies to a discrete random variable. However, without any loss of generality, the theorem is valid for a continuous random variable also. There, instead of a summation over some finite values, an integration over the domain of the random variable has to be performed.

(iv) The mathematical expectation of the sum of a given number of random variables is the sum of their expectations. If X and Y are two random variables, the mathematical expectation of $X + Y$ is

$$E(X + Y) = E(X) + E(Y)$$

(v) The mathematical expectation of the product of a given number of *independent random variables* is the product of their expectations. If X

and Y are two independent random variables, the mathematical expectation of XY is

$$E(XY) = E(X) \cdot E(Y)$$

The *variance* of a random variable X is given by

$$V(X) = E[X - E(X)]^2 = E(X^2) - [E(X)]^2$$

The variance of the random variable in Example 10.5 (tossing of a coin) can be computed in the following manner.

First we compute

$$E(X^2) = 20^2 \cdot \frac{1}{2} + 10^2 \cdot \frac{1}{2} = 200 + 50 = 250$$

$$[E(X)]^2 = 5^2 = 25$$

Now

$$V(X) = \sigma_X^{\ 2} = 250 - 25 = 225$$

Also the standard deviation of X

$$\sigma_X = \sqrt{225} = \text{Rs. } 15$$

10.4.2 Theorem on Variance

(i) The variance of a constant is zero. If c is a constant, then

$$V(c) = 0$$

(ii) The variance of the product of a constant and a random variable is the product of the square of the constant and the variance of the random variable. If c is a constant and X is a random variable, then

$$V(cX) = c^2 V(X)$$

(iii) The variance of the sum of a given number of random variables is the sum of their variances. If X and Y are two random variables, the variance of $X + Y$ is

$$V(X + Y) = V(X) + V(Y) + 2Cov(X, Y)$$

Here, $Cov(X, Y)$ is called the covariance between X and Y. We should note that covariance is a measure of simultaneous variability of the two variables.

Covariance can be shown as

$$Cov(X, Y) = E[\{X - E(X)\}\{Y - E(Y)\}] = E(XY) - E(X)E(Y)$$

But, if X and Y are independent, then a variation in one variable does not cause a variation in the other variable. Consequently, $Cov(X,Y) = 0$ and

$$V(X+Y) = V(X) + V(Y)$$

It may be noted here that all the theorems on mathematical expectation and variance discussed above are valid for both discrete and continuous random variables.

We can now state and prove an important result.

10.4.3 Standard Normal Variate

For any variable (random or otherwise) with a given mean and standard deviation, whenever the mean is subtracted from it and the result is divided by the standard deviation; the resultant variable has a mean equal to zero and a standard deviation equal to one.

Let us prove the above statement.

Let X be a random variable with mean (expectation) μ and standard deviation σ. Suppose

$$z = \frac{X-\mu}{\sigma}$$

$$E(z) = E\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma}E\left(X-\mu\right) = \frac{1}{\sigma}[E(X) - E(\mu)] = \frac{1}{\sigma}(\mu-\mu) = 0$$

Now

$$V(z) = V\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma^2}V\left(X-\mu\right) = \frac{1}{\sigma^2}[V(X) + (-1)V(\mu)]$$

The covariance term $Cov(X,\mu)$ above vanishes because X and μ are independent.

Thus

$$V(z) = \frac{1}{\sigma^2}[V(X)] = \frac{1}{\sigma^2}\sigma^2 = 1 \quad [\because V(\mu) = 0]$$

Now

$$Std.\,dev(z) = \sqrt{V(z)} = \sqrt{1} = 1$$

In this way we can see

$$E(z) = E\left(\frac{X-\mu}{\sigma}\right) = 0$$

and

$$Std.\,dev(z) = Std.\,dev\left(\frac{X-\mu}{\sigma}\right) = 1$$

In the next unit, we shall consider how this result is used in the context of the normal distribution.

The variable z defined in the above manner is called *standard normal variate.*

Check Your Progress 1

1. Check whether the following is a valid probability mass function or not.

$$p(x) = \frac{x^2 - x}{16}, \text{X} = -2,\ -1,\ 0,\ 1,\ 2,\ 3.$$

……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………

2. Find k such that the following is a valid probability mass function.

$$p(x) = \frac{k}{x^2}, \text{X} = 1, 2$$

……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………

3. A and B throw a dice for an amount of Rs. 99 to be won by one who throws a six first. If A has the first throw, what are their respective expected gains?

……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………

4. A contractor spends Rs. 3,000 to prepare for a bid on a construction project which, after deducting the manufacturing and the cost of bidding, will yield a profit of Rs. 25,000 if the bid is won. If the chance of winning the bid is ten percent, compute the contractor's expected profit and state the likely decision on whether to bid or not.

……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………

5. Prove the following results

   a) $E(cX) = cE(X)$, where c is a constant.

   b) $V(c) = 0$, where c is a constant.

   c) $V(cX) = c^2V(X)$, where c is a constant.

   ……………………………………………………………………………………

   ……………………………………………………………………………………

   ……………………………………………………………………………………

   ……………………………………………………………………………………

   ……………………………………………………………………………………

10.5 BINOMIAL DISTRIBUTION

The binomial distribution is an example of a discrete probability distribution. James Bernoulli presented it in the year 1700. The word 'binomial' suggests two. It signifies two possible outcomes of an experiment, the occurrence of an event or the non-occurrence of the event. A probability experiment can be termed as a Bernoulli experiment, if it satisfies the following conditions.

1. The experiment consists of a sequence of n repeated trials.
2. Each trial results in an outcome that may be classified either as a *success* or a *failure*.
3. The probability of a success, denoted by p, is known and remains the same in each trial. Consequently, the probability of a failure, denoted by $q = (1 - p)$ is also known and remains the same in each trial.
4. The trials are independent.

We may get some idea about a Bernoulli experiment by considering the experiment of tossing a coin a certain number of times and counting the number of heads appearing. Suppose, the coin is unbiased and we toss it 5 times. It is clear that the experiment consists of a sequence of 5 identical trials. There are two possible outcomes of each toss, a head (success) and a tail (failure). The probability of a head (success) is $\frac{1}{2}$ and this does not vary from one toss to another. The probability of a tail (failure) is again $\frac{1}{2}$ and this also does not vary from one toss to another. Finally, the tosses are independent in the sense that the outcome of one toss in no way depends upon the outcome of another toss. Thus, we find that this experiment of tossing a coin a certain number of times and observing the number of heads appearing satisfies all the conditions of a Bernoulli experiment.

In a Bernoulli experiment, we are interested in deriving the probability of a given number of successes, say, x occurring in n trials. (For example, in the previous example we may be interested in finding out the probability of getting 3 heads in 5 tosses.) It is clear that the random variable x can assume values 0, 1, 2, 3, …, n. Suppose, we denote a success by S and a failure by F, then x successes and $(n - x)$ failures may occur in a number of different sequences. One possible sequence is that the first x trials are all successes and the remaining $(n - x)$ trials are all failures. Symbolically the sequence is shown by

$$\frac{SS \dots . S}{x \text{ times}} \quad \frac{FF \dots . F}{(n-x) \text{ times}}$$

The probability of the above sequence of x successes and $(n - x)$ failures can be obtained by applying the multiplication theorem of probability. The probability is

$$\frac{pp \dots .. p}{x \text{ times}} \quad \frac{(1-p)(1-p) \dots .. (1-p)}{(n-x) \text{ times}} = p^x (1-p)^{n-x}$$

But, as mentioned earlier, x successes and $(n - x)$ failures can occur in other sequences also. However, each of the sequences in which x successes and $(n - x)$ failures occur will have a probability of $p^x (1-p)^{n-x}$. Thus, the probability of x successes in n trials is the probability of the occurrence of the x successes and $(n - x)$ failures in any of the possible sequences. This probability can be obtained by applying the addition theorem of probability over the possible sequences. But, as the probability of x successes and $(n - x)$ failures is the same for each of the possible sequences, the required probability of x successes in n trials is the product of the total number of possible sequences and the probability of the occurrence of a sequence. The total number of sequences in which x successes (and $n - x$ failures) can occur in n trials is basically a problem of obtaining the number of combinations of n things taken x at a time and is denoted by ${}^nC_x$. From the mathematics of *permutation* and *combination*, we have

$${}^nC_x = \frac{n!}{x!(n-x)!}$$

where

$$n! = n(n-1)(n-2) \dots 2.1$$

$$x! = x(x-1)(x-2) \dots 2.1$$

and $0! = 1$.

We should note that the notation '!' is called *factorial*.

For example, $4! = 4 \times 3 \times 2 \times 1 = 24$. The symbol C in ${}^nC_x$ represents combination.

For example, ${}^5C_3 = \frac{5!}{3!(5-3)!} = \frac{5!}{3!2!}$

$$= \frac{5 \times 4 \times 3 \times 2 \times 1}{3 \times 2 \times 1 \times 2 \times 1}$$

$$= 10$$

Thus, the probability of x successes in n trials is given by

$$p(x) = {}^{n}C_{x}\, p^{x} (1-p)^{n-x}$$

$x = 0, 1, 2, \ldots\ldots, \text{n}.$

The above expression is the probability mass function for the Binomial distribution. This function has been used for presenting the Binomial distribution of x = 0, 1, 2, … n successes in n trials in Table 10.3. We note that a binomial distribution has two parameters n and p. It means that the distribution is completely specified if the values of n and p are known.

Table 10.3: Binomial Distribution

Number of Successes x	Probability $p(x)$
0	$(1-p)^{n}$
1	$np(1-p)^{n-1}$
2	$\frac{n(n-1)}{2.1} p^{2} (1-p)^{n-2}$
$\vdots$	$\vdots$
N	p^{n}
Total	**1**

Now let us find out the mean of the binomial distribution.

Let there be n number of trials in a Bernoulli experiment with p as the probability of a success in a trial. This implies the probability of failure is ***q***

Mean

$$E(X) = \sum_{x=0}^{n} \text{x}\ {}^{n}C_{x}\, p^{x} (1-p)^{n-x}$$

By simplifying the above we find that the mean of binomial distribution is given by np. We do not present the proof of the above as it is quite cumbersome.Similarly the variance of binomial distribution is given by

$$V(X) = E[X - E(X)]^{2} = E(X^{2}) - [E(X)]^{2}$$

By simplification of the above it can be shown that the variance of binomial distribution is given by npq[which is equal to $np(1\text{-}p)$].

Thus we observe that the mean and variance of binomial distribution are given by its two parameters *n* and *p*. We give some examples on the applications of binomial distribution below.

Example 10.6

A machine is generally known to be producing 20 percent defective items. A quality control Inspector selects 5 items at random. Find the probability of getting (i) exactly 1 defective item (ii) at least 3 defective items.

This is an example of binomial distribution with $p = 0.20$ and $n = 5$. Let us now solve the question.

(i) We know that the probability of x defective items (i.e., $n - x$ non-defective items) in n items is ${}^{n}C_{x}\,p^{x}(1-p)^{n-x}$. Here $n = 5$ and $x = 1$. Thus the probability of exactly 1 defective item is

$p(1) = {}^{5}C_{1}(0.20).(0.80)^{4} = 0.4096$.

(ii) At least 3 defective items means that there can be 3 or 4 or 5 defective items. Thus the probability of at least 3 defective items is the probability of 3 defective items plus the probability of 4 defective items plus the probability of 5 defective items.

Now, the probability of 3 defective items is

$p(3) = {}^{5}C_{3}(0.20)^{3}(0.80)^{2} = 0.0512$

The probability of 4 defective items is

$p(4) = {}^{5}C_{4}(0.20)^{4} \cdot 0.8 = 0.0064$

The probability of 5 defective items is

$P(5) = {}^{5}c_{5}(0.20)^{5} = 0.0003$

Therefore, the probability of at least 3 defective items = 0.0512 + 0.0064 + 0.0003 = 0.0579.

Example 10.7

If a fair dice is thrown 36 times, what is the expected number of times of getting a 6? What is the variance?

If p is the probability of getting a 6, then $p = \frac{1}{6}$ and $(1-p) = \frac{5}{6}$. Now, $n = 36$.

The mathematical expectation

$np = 36 \cdot \frac{1}{6} = 6$

Thus, one can expect to get a 6, six times when a dice is thrown 36 times.

The variance is

$$np(1-p) = 36 \cdot \frac{1}{6} \cdot \frac{5}{6} = 5$$

10.6 POISSON DISTRIBUTION

The Poisson distribution is another discrete probability distribution. It is named after a French mathematician Simeon Poisson who derived this distribution in 1837. The distribution is in fact a special (limiting) case of binomial distribution. When the probability of success, p, in a binomial distribution is very small and the number of trials, n, is so large that the expectation, $\mu = np$, is a finite quantity; the binomial distribution tends to Poisson distribution. This distribution is particularly useful while dealing with the number of occurrences of something over a specified interval of time or space. For example, the random variable under consideration may be the number of telephone calls arriving at a telephone switch-board in 1 hour, the number of leaks in 100 kilometres of pipeline, or the number of bus accidents reported in a particular day in Delhi.

To obtain the probability mass function of Poisson distribution, we can consider the example of the number of telephone call, x, in an hour and assume that the expected number of telephone calls per hour (i.e., the mathematical expectation) is λ For the applicability of the binomial distribution, we divide the interval of one hour into sub-intervals that are so small that the probability p of having a telephone call in a sub-interval is very low and that of getting more than one call is approximately zero. Thus, each sub-interval can be treated as a Bernoulli trial having only two possible outcomes; either there will be a telephone call (a success) or no telephone call (a failure). The number of sub-intervals is taken to be equal to n, the total number of trials. We note that the expected number of telephone calls, λ, remains the same and is equal to np (as we have seen from the binomial distribution). Therefore, the probability of a telephone call in a sub-interval is $\frac{\lambda}{n}$.

Thus, the probability of x telephone calls in one hour amounts to finding the probability of x successes in n trials when n tends to infinity (as argued above, n trials correspond to n sub-periods that constitute one hour and n tends to infinity as a result of making each sub-trial extremely small so that the total number of trials tends to be infinitely large). This probability is given as the following limit of a binomial distribution.

$$\lim_{n \to \infty} {}^{n}C_{x} \left(\frac{\lambda}{n}\right)^{x} \left(1 - \frac{\lambda}{n}\right)^{n-x}$$

Let us try to find the above-mentioned limit.

The Poisson probability mass function is derived from the above and is given by

$$p(x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

$$x = 0, 1, 2, \ldots$$

where, X is a random variable denoting the number of successes in a specified time interval or length interval.

λ = expected value or average number of occurrences in an interval of time or length etc.

e = a constant (base of the natural logarithm) whose value is $e = 2.7182$.

Table 10.4: Poisson Distribution

The value of the Poisson random variable x	Probability
0	$e^{-\lambda}$
1	$\frac{\lambda e^{-\lambda}}{1!}$
2	$\frac{\lambda^2 e^{-\lambda}}{2!}$
.	.
.	.
Total	**1**

In Poisson distribution, there is no upper limit on the random variable x, the number of occurrences. It is a discrete random variable that can assume an infinite sequence of values (x = 0, 1, 2,3, …). The distribution has only one parameter λ. Table 10.4 presents the Poisson distribution generated by the Poisson probability mass function.

Expectation and Variance

The expectation of the Poisson distribution is given by the constant λ. It can be shown that the variance of the Poisson distribution is also given by λ.

Example 10.8

An analysis shows that on an average 10 cars arrive at a petrol pump in a 15-minute period of time.

i) Find the probability of the arrival of 5 cars in 15 minutes.

ii) What is the probability of the arrival of 1 car in 3 minutes?

Here, λ = 10 and x = 5. Thus, the required probability is

$$p(5)=\frac{10^5 e^{-10}}{5!}=0.0378.$$

Since the expected number of arrivals in a 15-minute period is 10, the expected number of arrivals in a 3-minute period is $\frac{10}{15}\times 3=2$. Thus for the second part of the question, λ = 2 and x = 1. Hence, the probability of one arrival in a 3-minute period is

$$p(1)=\frac{2e^{-2}}{1!}=0.2707.$$

Check your progress 2

1. In a hospital, there are 3 ambulances for the transportation of patients. The probability of the availability of an ambulance is 0.75. If an ambulance is needed what is the probability that

   a) no ambulance will be available?

   b) at least one ambulance will be available?

   ……………………………………………………………………………………
   ……………………………………………………………………………………
   ……………………………………………………………………………………
   ……………………………………………………………………………………
   ……………………………………………………………………………………

2. Can for a binomial distribution the mean and the variance be 3 and 5 respectively?

   ……………………………………………………………………………………
   ……………………………………………………………………………………
   ……………………………………………………………………………………
   ……………………………………………………………………………………
   ……………………………………………………………………………………

3. It is known from the past experience that in a certain plant, there are on an average 4 industrial accidents per month. Find the probability that in a given month, there will be less than 4 accidents. Assume Poisson distribution.

   ……………………………………………………………………………………
   ……………………………………………………………………………………
   ……………………………………………………………………………………
   ……………………………………………………………………………………
   ……………………………………………………………………………………

10.7 LET US SUM UP

In this unit, we have used the concept of probability to understand the meaning of a probability distribution. We have understood the concepts of the mathematical expectation and the variance of a probability distribution. We have distinguished between a discrete probability distribution and a continuous probability distribution. In this context, we have introduced the notions of a probability mass function and a probability density function. We have studied two specific discrete probability distributions. They are the binomial distribution and the Poisson distribution. We have learnt the characteristics of these distributions, particularly, the expressions for their mean and variance. We have tried to understand the use of these two distributions in various situations.

10.8 ANSWERS OR HINTS TO CHECK YOUR PROGRESS EXERCISES

Check Your Progress 1

1) $p(-2) = \frac{3}{8}, p(-1) = \frac{1}{8},$

$p(0) = 0$

$p(1) = 0$

$p(2) = \frac{1}{8}$

$p(3) = \frac{3}{8}$

Since p(x) $\geq$ 0 for all values of x, the first condition is satisfied.

$\Sigma\, p(x) = \frac{3}{8} + \frac{1}{8} + 0 + 0 + \frac{1}{8} + \frac{3}{8} = 1$.

Therefore, second condition is also satisfied.

Hence, the given function is a valid probability mass function.

2) $k = \frac{4}{5}$

3) *A* can win if *A* throws a six in the first throw *or A* cannot throw a six in the first throw *and* B cannot throw a six in the second throw *and* A throws a six in the third throw, and so on. Therefore, the probability that *A* throws a six first

$$= \frac{1}{6} + \frac{5}{6}\cdot\frac{5}{6}\cdot\frac{1}{6} + \frac{5}{6}\cdot\frac{5}{6}\cdot\frac{5}{6}\cdot\frac{5}{6}\cdot\frac{1}{6} + \ldots = \frac{6}{11}$$

Thus, the expected gain of *A* is $99\cdot\frac{6}{11}$ = Rs. 54

The probability that B throws a six first

$= 1 - P\,(A \text{ throws a six first}) = 1 - \frac{6}{11} = \frac{5}{11}$

So, the expected gain of B is $99 \cdot \frac{5}{11}$ = Rs. 45

4) Rs. (-)200

5) a) $E\,(cX) = E\,(c)\,E\,(X) = cE\,(X)$ (because, expectation of a constant is the constant itself)

b) $V\,(c) = E\left(c^2\right) - \left[E\,(c)\right]^2 = c^2 - c^2 = 0$

c) $V\,(cX) = E\,\left(c^2X^2\right) - \left[E\,(cX)\right]^2 = c^2E\,\left(X^2\right) - c^2\left[E\,(X)\right]^2$

$= c^2\left[E\,\left(X^2\right) - \left\{E\,(X)\right\}^2\right] = c^2V\,(X)$

Check Your Progress 2

1) a) $\frac{1}{64}$ b) $\frac{63}{64}$

2) Mean $np = 3$, variance $np(1 - p) = 5$. Now,

$$\frac{np(1-p)}{np} = \frac{5}{3}$$

$$or\ 1 - p = \frac{5}{3}$$

$$or\ \text{q} = 1 - \frac{5}{3} = -\frac{2}{3} < 0, \text{ which is not possible.}$$

3) 0.4332.

UNIT 11 CONTINUOUS PROBABILITY DISTRIBUTIONS*

Structure

11.0 OBJECTIVES

After going through this unit you should be able to

- explain and use the normal distribution;
- explain the concept of the degrees of freedom; and
- form some elementary ideas about the chi-squared distribution, the student's-*t* distribution and the *F* distribution.

11.1 INTRODUCTION

In the previous Unit, we made a distinction between a discrete random variable and a continuous random variable. In that unit, we introduced the concept of a probability distribution and found that it is essentially a statement regarding the values taken by a random variable with their associated probabilities. We studied two important discrete probability distributions namely, the binomial distribution and the Poisson distribution. In the present Unit we will continue with the topic and study a very important continuous probability distribution called the normal distribution. It may be mentioned that the normal distribution plays an important role in the statistical inference and tests of hypotheses and these are going to be our subject matter of Block-7.

In fact, we will consider in Unit 13 the topic of sampling distribution that forms the foundation of statistical inferences and tests of hypotheses. However,

* Dr. Anup Chatterjee (retd.), ARSD College, University of Delhi

ignou THE PEOPLE'S UNIVERSITY

sampling distribution can be properly appreciated only if we have some rudimentary ideas about three other continuous probability distributions besides the normal distribution, viz., the chi-squared distribution, the student's-*t* distribution and the *F* distribution. We will discuss about these probability distributions below.

11.2 NORMAL DISTRIBUTION

Normal distribution is perhaps the most widely used distribution in Statistics and related subjects. It has found applications in inquiries concerning heights and weights of people, IQ scores, errors in measurement, rainfall studies and so on. Abraham de Moivre gave the mathematical equation for the normal distribution in 1733. Karl Friedrich Gauss also independently derived its equation from a study of errors in repeated measurements of the same quantity. Accordingly, sometimes it is also referred to as the Gaussian distribution. The distribution has provided the foundation for much of the subsequent development of mathematical statistics.

We have seen in the previous Unit that for a continuous random variable, the counterpart of a probability mass function is the *probability density function.* We shall denote the probability density function also by $p(x)$. The probability density function of a continuous random variable that follows the normal distribution is given by

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

where $-\infty < x < \infty$, and

$\pi = 3.17141$ (approximately)

$e = 2.71828$ (approximately).

It is clear that the normal density function is completely determined by the parameters μ and σ. It means that given the values of μ and σ, we can trace out the normal curve by obtaining the values of $p(x)$ for different values of x. In fact, it can be shown that μ and σ are respectively the mean and the standard deviation of the normal distribution. When a random variable X follows normal distribution with mean μ and standard deviation σ we write it in symbols as $X \sim N(\mu, \sigma)$ and read as '*X* follows normal distribution with mean μ and standard deviation σ. The normal curve is a symmetrical bell-shaped curve as shown in Fig. 11.1.

The important features of the normal distribution are as follows:

1) The normal curve stretches from $-\infty$ to $+\infty$. This means that a normal random variable (X) assumes values between $-\infty$ to $+\infty$.

2) The curve is symmetric about its mean, i.e., $\bar{x} = \mu$. This means that corresponding to $x = \mu + a$ and $x = \mu - a$, the values of $p(x)$ are the same (for any arbitrarily chosen 'a').

3) The median and the mode of the distribution coincide with the mean. Thus mean = median = mode = μ.

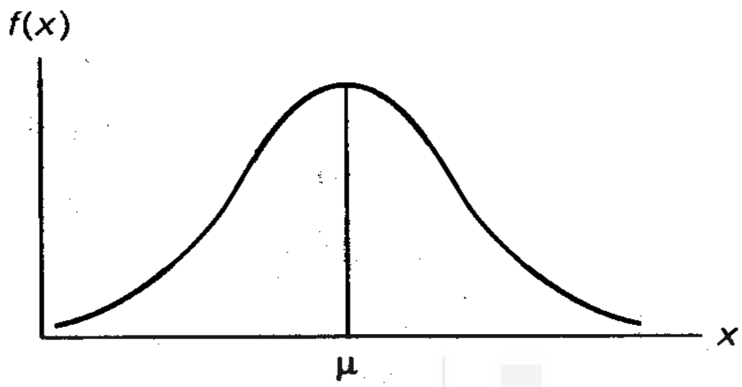

Fig 11.1: Normal Curve

4) The maximum of the normal curve occurs at $x = \mu$. Thus $p(x)$ is maximum when x= μ.

5) The points of inflexion of the normal curve occurs at $x = \mu + \sigma$ and $x = \mu - \sigma$. At the points of inflexion, the normal curve changes its curvature.

The following area-properties hold for a normal distribution. In Fig.11.2 below we plot a normal curve with mean $\mu = 50$ and standard deviation $\sigma = 4$.

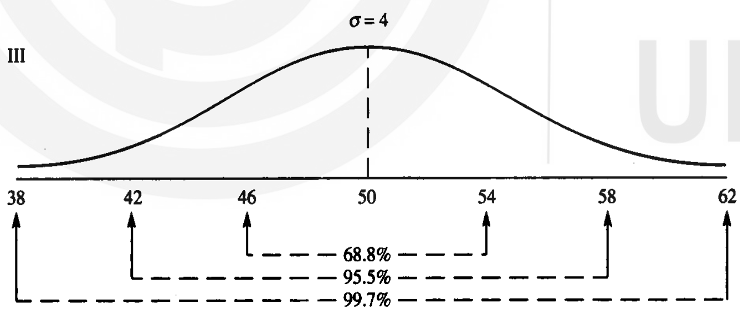

Fig. 11.2: Area under Normal Curve

(a) 68.8 % of the area under the normal curve lies between the ordinates at $\mu - \sigma$ and $\mu + \sigma$. Thus in Fig.11.2, 68.8% area is covered when x ranges between 46 and 54.

(b) 95.5% of the area under the normal curve lies between the ordinates at $\mu - 2\sigma$ and $\mu + 2\sigma$. In Fig. 11.2, 95.5% area is covered when $42 \leq X \leq 58$.

(c) 99.7% of the area (i.e., almost the whole of the distribution) under the normal curve lies between the ordinates at $\mu - 3\sigma$ and $\mu + 3\sigma$. In Fig.11.2 we find that 99.7% area is covered when $38 \le X \le 62$.

If we have different values of μ and σ, the range of x mentioned in Fig. 11.2 will change.

11.2.1 Standard Normal Curve

We have seen in the previous unit that the curve for any continuous probability distribution or probability density function is so traced out that the area under the curve bounded by the two ordinates corresponding to $x = x_1$ and $x = x_2$ gives the probability that the random variable assumes a value between $x = x_1$ and $x = x_2$. Thus, for a normal curve

$$P(x_1 \le X \le x_2) = \int_{x_1}^{x_2} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \frac{1}{\sqrt{2\pi}\sigma} \int_{x_1}^{x_2} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

Obviously, this probability depends upon the values of the two parameters μ and σ. However, it is very difficult to solve the above-mentioned integral of the normal distribution. This has necessitated the tabulation of normal curve areas for quickly obtaining the probabilities of the normal variable assuming values in different intervals. But, it is really meaningless to attempt to construct a separate table for every conceivable combination of values for μ and σ. Fortunately, the solution to an apparently hopeless task has been achieved by the application of a standard result in statistics that we have seen and proved in the last unit. Let us recapitulate the result. We have seen that

For any variable with a given mean and standard deviation, whenever the mean is subtracted from the variable and the result is divided by the standard deviation; the resultant variable has a mean equal to zero and a standard deviation equal to one.

Thus if X is a variable with mean (expectation) μ and standard deviation σ then $z = \frac{X-\mu}{\sigma}$ has a mean equal to zero and standard deviation equal to one.

It means that normal variables with different combinations of μ and σ can all be transformed into a unique normal variable with mean 0 and standard deviation 1.

Thus if X is a normal variable with mean (expectation) μ and standard deviation σ, then $z = \frac{X-\mu}{\sigma}$ for any combination of μ and σ, is always a normal variable with mean 0 and standard deviation 1.

Symbolically,

If $X \sim N(\mu,\sigma)$

then,

$$z = \frac{X-\mu}{\sigma} \sim N(0,1)$$

for any μ and σ.

Such a transformed normal variable is called a *standard normal variate*. The probability density function of the standard normal variate z is given by

$$p(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \qquad -\infty < z < \infty$$

Once we obtain a standard normal variate, our seemingly hopeless task of obtaining probability areas for different combinations of μ and σ becomes elegantly simple. Let us see how. We should note that a standard normal variate has a unique mean of 0 and a unique standard deviation of 1. It means, if we can construct a table for probability areas of such a unique standard normal variate, it can be used for obtaining probability for any normal variable with any combination of mean and standard deviation. The only thing is that the given normal variable is to be transformed into the standard normal variate. In fact, such a table for areas (or probability) has been compiled for a standard normal variate (see Appendix Table A1: Area under the Standard Normal Curve given at the end of this book) and is very much in use in statistics. Thus, for the computation of the required probability for any normal variable with some mean and standard deviation, the upper and the lower limits say, $x = x_1$ and $x = x_2$ of the given interval are converted into the corresponding z-values say, $z = z_1$ and $z = z_2$ and the relevant area is obtained from Appendix Table A1 given at the end of this book.

Remember that the standard normal curve is symmetrical and it covers an area of 1.0 Since the value of z ranges between $-\infty$ and ∞, we find that the area between $0 < z < \infty$ is 0.5 (half the area under standard normal curve). Similarly, the area between $-\infty < z < 0$ is 0.5. Since the standard normal curve is symmetric we have one advantage; the area under the curve is the same on both sides. In Appendix Table A1 the area for different positive values of z are given.

If we look into column 1 of Appendix Table A1 we find that values assumed by z ranges from 0.0 to 3.0. Corresponding to each value there are 10 columns marked .00, .01, ..., .09. These columns represent the second digit after decimal. For example, if $z = 0.45$, then we look for the row corresponding to 0.4. On this row we move to the right and look for the column representing .05. In Table A1 we find that when z=0.45 the area covered is 0.1736. Note that when $z = -0.45$ the area under standard normal curve again is 0.1736. As another example, the area under the standard normal curve when $z = 1.31$ is 0.4049.

Theoretically z can assume any value between $-\infty$ and ∞. However, when z = 3.09 the area covered is 0.4990. Therefore, in Table A1 areas for $z > 3.09$ are not given.

Let us now consider some examples to see the applications of the normal area table.

Example 11.1

Find the area under the standard normal curve in each of the following cases by using Appendix Table A1 for areas under the standard normal curve.

a) Between $z = 0$ and $z = 1.8$.

b) Between $z = -0.25$ and $z = 0$.

c) Between $z = -0.52$ and $z = 2.25$.

Solution:

a) In Table A1, let us move downward under the column marked z until we reach the entry 1.8. Now, let us turn right to the column marked 0. We find here an entry equal to 0.4641. This is the required area.

b) Since the standard normal curve is symmetric about the mean, the required probability between $z = -0.25$ and $z = 0$ can be obtained by finding the area between $z = 0$ and $z = 0.25$ from the table. So, let us move downward under the column marked z until we reach the entry 0.2. Then we turn right to the column marked .05. We find here an entry equal to 0.0987. This is the required area.

c) It is clear that the required area is

(area between $z = -0.52$ and $z = 0$) + (area between $z = 0$ and $z = 2.25$)

= (area between $z = 0$ and $z = 0.52$) (by symmetry) + (area between $z = 0$ and $z = 2.25$)

= 0.1985 + 0.4878

= 0.6863.

Example 11.2

In a sample of 120 workers in a factory the mean and standard deviation of daily wages are Rs. 11.35 and Rs. 3.03 respectively. Find the percentage of workers getting wages between Rs. 9 and Rs. 17 in the whole factory assuming that the wages are normally distributed.

Solution: Let x be a random variable denoting wages. Then, x is a normal variable with mean $\mu = 11.35$ and the standard deviation $\sigma = 3.03$. The corresponding standard normal variate

$$z = \frac{x - 11.35}{3.03}$$

For $x = 9,\ z = \dfrac{9 - 11.35}{3.03} = \dfrac{-2.35}{3.03} = -0.78$

For $x = 17,\ z = \dfrac{17 - 11.35}{3.03} = \dfrac{5.65}{3.03} = 1.86$

$\therefore\ P(9 \le x \le 17) = P(-0.78 \le z \le 1.86)$

$= P(-0.78 \le z \le 0) + P(0 \le z \le 1.86)$

$= P(0 \le z \le 0.78) + P(0 \le z \le 1.86)$

$= 0.2823 + 0.4686$

$= 0.7509.$

Thus, 75.09% of workers get wages between Rs. 9 and Rs. 17.

11.2.2 Normal Approximation to the Binomial Distribution

Sometimes in statistics, one distribution is obtained as the limiting form of another distribution. For example, in Unit 10, we have learnt that at times the probability of success, p, in a binomial distribution is very small and the number of trials, n, is so large that the expectation, $\mu = np$, is a finite quantity. In such cases the binomial distribution tends to Poisson distribution. You can recall that both the binomial distribution and the Poisson distribution are discrete distributions. However, the limiting form of the discrete binomial distribution need not be uniquely the discrete Poisson distribution. In fact, when n is very large and p is not extremely close to 0 or 1; the binomial distribution approaches the continuous normal distribution. As a result, in real world situations, the normal distribution is often used for approximating the binomial distribution. We have already seen that the standard normal table comes very handy in obtaining probabilities for a normal variable falling within a specified interval. We may now state a result that helps us to use areas under the standard normal curve to approximate the binomial properties of a random variable:

If X is a binomial variable with mean $\mu = np$ and variance $\sigma^2 = npq$ then

$$z = \frac{X - np}{\sqrt{npq}}$$

tends to a normal distribution with mean zero and standard deviation one, as n tends to infinity
In symbols,

$$\lim_{n\to\infty} \frac{X - np}{\sqrt{npq}} \sim N(0,1)$$

It has been observed that the normal distribution provides a good approximation for a binomial distribution even when n is not very large and p is reasonably close to 0.5.

Let us consider an example.

Example 11.3

Suppose the probability of a particular kind of machine being defective is 0.4. A quality control inspector examines a lot of 15 such machines to identify the defective machines. Find the probability that the inspector will find 4 machines to be defective.

Solution: We know from our discussion on binomial distribution that the distribution of defective machines in a given lot is a binomial variable. Here

$n = 15$, $p = 0.4$ and $q = 1 - p = 0.6$

The probability of 4 defective machines is given by

$$^{15}c_4 (0.4)^4 (0.6)^{11} = \frac{15!}{4!11!}(0.4)^4 (0.6)^{11} = 1365 \times 0.0256 \times 0.0036279 = 0.1268$$

Now suppose, we want to find the required probability by the normal approximation. Then

$np = 15 \times 0.4 = 6$, $npq = np(1-p) = 15 \times 0.4 \times 0.6 = 3.6$ and $\sqrt{npq} = \sqrt{3.6} = 1.897$

We have

$$z = \frac{X - np}{\sqrt{npq}}$$

which is approximately a standard normal variate. But the standard normal variate is a continuous random variable. We know that for such a random variable, the probability of its taking a particular value cannot be determined. It is only the probability of the random variable lying within an interval that can be obtained. Thus the probability of the binomial variable taking a value 4 has to be translated into the probability of the corresponding normal variable falling in an interval around the value 4 for the required normal approximation. Since, a binomial variable assumes zero and positive integer values; the value immediately preceding 4 and the value immediately succeeding 4 that the binomial variable under question can take are 3 and 5 respectively. As a result, the probability of the binomial variable taking a value equal to 4 can be reasonably approximated by the probability of the corresponding normal variable falling within an interval (3.5, 4.5). The required probability is thus approximately equal to the area under the normal curve between the ordinates x_1 = 3.5 and x_2 = 4.5. Converting to z-values, we have

$$z_1 = \frac{x_1 - np}{\sqrt{npq}} = \frac{3.5 - 6}{1.897} = -1.32 \text{ and } z_2 = \frac{x_2 - np}{\sqrt{npq}} = \frac{4.5 - 6}{1.897} = -0.79.$$

If X is the binomial variable and z is the corresponding standard normal variate, then $P(X = 4) = P(-1.32 \leq z \leq -0.79) = 0.1214$ (From the area under standard normal curve given in Appendix Table A1)

We can see that the normal approximation of 0.1214 agrees quite closely with the actual probability of 0.1268 for 4 defective machines obtained from the binomial distribution.

Check Your Progress 1

1) Find the area under the normal curve in the following cases.
   a) Between $z = 1.55$ and $z = 2.55$.
   b) To the left of $z = -1.5$.
   c) To the right of $z = 2.5$.

   ……………………………………………………………………………………

   ……………………………………………………………………………………

   ……………………………………………………………………………………

   ……………………………………………………………………………………

2) The mean height of 1000 men is 68 inches and the standard deviation is 5 inches. If the heights are normally distributed, find how many men have heights between 67 inches and 69 inches.

   ……………………………………………………………………………………

   ……………………………………………………………………………………

   ……………………………………………………………………………………

   ……………………………………………………………………………………

3) A company, that sells 5000 batteries in a year, guarantees them for a life of 24 months. The life of the batteries is estimated to be approximately normal with mean equal to 34 months and standard deviation equal to 5 months. Find the number of batteries that will have to be replaced under the guarantee.

   ……………………………………………………………………………………

   ……………………………………………………………………………………

   ……………………………………………………………………………………

   ……………………………………………………………………………………

11.3 SOME OTHER CONTINUOUS DISTRIBUTIONS

There are some other continuous probability distributions that play important roles in various branches of statistics. In Block 4, we will study statistical inference. In that Block, in addition to normal distribution, we will often use concepts of three more continuous distributions, C*hi-Squared* (Pronounced as *kai-squared* and denoted by the symbol χ^2), the *Student's-t distribution* and the *F-distribution.* In this section, we are going to discuss these distributions in brief. We begin with a general concept, the 'degrees of freedom', which finds applications in all these distributions.

11.3.1 Degrees of Freedom

In connection with these distributions, we shall often come across a term: degrees of freedom. Let us get some idea about the term now. In a general sense, *the degrees of freedom refers to the number of pieces of independent information* that are required to compute some characteristic (say, variance) of a given set of observations. We consider an example here.

Suppose there are 5 observations: 4, 7, 12, 3 and 15. Hence, the arithmetic mean $\overline{X}$ is 8. The computation of variance $\frac{1}{n}\sum(x_i - \overline{x})^2$ involves obtaining the squares of the deviations of the values of the observations from their arithmetic mean and adding them as shown below:

$$(4-8)^2 + (7-8)^2 + (12-8)^2 + (3-8)^2 + (14-8)^2$$
$$= (-4)^2 + (-1)^2 + (4)^2 + (-5)^2 + (6)^2$$

From the properties of arithmetic mean ($\overline{X}$), we know that the summation of the values inside the brackets must be equal to zero, i.e., in general $\sum_{i=1}^{n}(x_i - \overline{X}) = 0$. It means that in the computation of the variance, if the first four terms inside the brackets happen to be –4, –1, 4 and –5 respectively, the fifth term cannot be any other term but 6. Thus in this example, we do not have 5 independent pieces of information inside the brackets. The fifth piece of information inside the bracket, i.e., '6' depends upon the first four pieces of information inside their respective brackets.

We can fix the idea better if we think of a person who does not have any idea about the individual observations. She is only told that there are 5 observations and the first 4 deviations of the values from the mean of the five observations are –4, –1, 4 and –5 respectively. She is then asked to calculate the variance of the 5 observations. If she knows the law that the sum of the deviations of the values of a variable from their arithmetic mean is always equal to zero, she will at once be able to figure out that the deviation of the 5th value from its arithmetic mean is 6. She will then proceed to take the squares of these deviations and add them together to arrive at the last step for the computation of the required variance. The last step involves a division by the number of observations to get an idea about the average dispersion of the values about their arithmetic mean (we take variance as the required measure here). What is the appropriate value that she should take for the number of observations? Is it 5? Let us probe a little into it. We have seen that the fifth deviation is determined by the first four deviations, which means that there are 4 independent pieces of information that produce the dispersions of the given 5 values about their arithmetic mean. Therefore, for measuring the average dispersion, i.e., the variance, the sum of the squares of the five deviations should be divided by 4 and not by 5.

Thus, in this example, the degrees of freedom are 4. Generalising the above example, for obtaining the variance of n observations, there are n–1 degrees of freedom because of the restriction $\sum_{i=1}^{N}(x_i - \overline{X}) = 0$. As a result, for variance, $\sum_{i=1}^{N}(x_i - \overline{X})^2$ is divided by n–1, i.e., $S^2 = \frac{1}{n-1}\sum_{i=1}^{n-1}(x_i - \overline{X})^2$, where, S^2 is the variance. From the above discussion, we can say that the degrees of freedom that we have for the computation of any characteristic is equal to the number of observations minus the number of restrictions put on the computation of the required characteristic. In symbols, *d.f.* = $n - r$ where, *d.f.* is the degrees of freedom, n is the number of observations and r is the number of restrictions.

We should note here that when n is quite large, for calculating the variance or its positive square root, i.e., the standard deviation, we often divide $\sum_{i=1}^{N}(x_i - \overline{X})^2$ by n and not by $n - 1$. However, strictly speaking, we should divide $\sum_{i=1}^{N}(x_i - \overline{X})^2$ by $n - 1$, particularly when n is small.

11.3.2 The χ^2 (Chi-Squared) Distribution

Suppose, X is a normal variable with mean (expectation μ and standard deviationσ, than $z = \frac{X-\mu}{\sigma}$ is a standard normal variate, i.e., $z \sim \mathrm{N}(0,1)$ If we take the square of z, i.e., $z^2 = \left(\frac{\mathrm{X}-\mu}{\sigma}\right)^2$, then z^2 is said to be distributed as a χ^2 variable with one degree of freedom and expressed as χ_1^2.

It is clear that since χ_1^2 is a squared term; for z laying between $-\infty$ and $+\infty$, χ_1^2 will lie between 0 and $+\infty$ (because a squared term cannot take negative values). Again since, z has a mean equal to zero, most of the values taken by z will be close to zero. As a result, the probability density of χ^2 variable will be maximum near zero.

Generalizing the result mentioned above, if $z_1, z_2, \ldots, z_k$ are independent standard normal variates (i.e., normal variables with zero mean and unit variance), then the variable

$$z = \sum_{i=1}^{ki} z_1^2$$

is said to be a χ^2 variable with k degrees of freedom and is denoted by χ_k^2 Fig, 15.3 given below shows the probability curves for χ^2 variables with different degree of freedom.

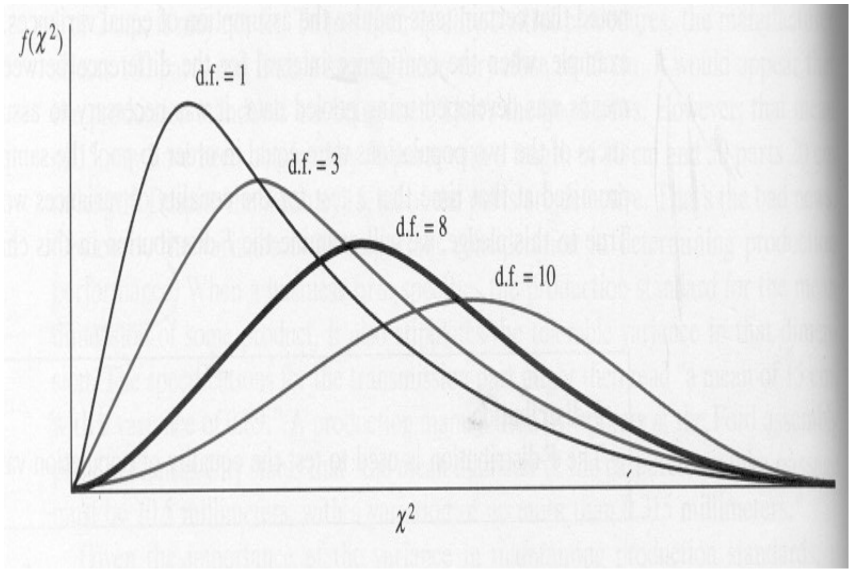

Fig 11.3: Chi-squared Probability Curves

We should note the following features of the χ^2 distribution.

1) As Fig. 11.3 shows, the χ^2 is a positively skewed distribution. Its degree of skewness depends on its degrees of freedom. For lower degrees of freedom, the distribution is highly skewed. As the number of degrees of freedom increases, the distribution becomes increasingly symmetric. In fact, for degrees of freedom more than 100, the variable $\sqrt{2\chi^2} - \sqrt{(2k-1)}$ can be treated as a standard normal variate, where *k* is the degrees of freedom.

2) The mean of the chi-squared distribution is *k*, and its variance is *2k*, where *k* is the degrees of freedom.

3) If Z_1 and Z_2 are two independent chi-squared variables with k_1 and k_2 degrees of freedom respectively, then $Z_1 + Z_2$ is also a chi-squared variable with degrees of freedom equal to $k_1 + k_2$.

As in the case of the normal distribution, a similar table has been prepared for the χ^2 distribution also (see Appendix Table A2 at the end of this book). We have to just consult this table to obtain the required probability of the χ^2 variable for different degrees of freedom.

In Table A2 *df* represent degrees of freedom. While the columns $\chi^2_{0.05}$ and $\chi^2_{0.01}$ denotes χ^2 values for 5% and 1% *level of significance* respectively. We will explain the concept of level of significance in Unit 14.

The following example illustrates the use of the chi-squared table.

Example 11.4

What is the probability of obtaining a χ^2 value of 34 or greater, given the degrees of freedom 25?

Solution: We can see from Appendix Table A2 that if we move down the degrees of freedom column to reach the figure of 25, the nearest figure to 34 that we find across the corresponding row is 34.08747. The probability for 34.08704, as we can see from the table, is 0.10. Hence the required probability is 0.10.

11.3.3 The Student's-*t* Distribution

W.S. Gosset presented the *t*-distribution. The interesting story is that Gosset was employed in a brewery in Ireland. The rules of the company did not permit any employee to publish any research finding independently. So, Gosset adopted the pen-name 'student' and published his findings about this distribution anonymously. Since then, the distribution has come to be known as the student's-*t* distribution or simply, the *t* distribution.

If z_1 is a standard normal variate, i.e., $z_1 \sim N(0,1)$ and z_2 is another independent variable that follows the chi-square distribution with k degrees of freedom, i.e.,

$z_2 \sim \chi_k^2$, then the variable $t = \frac{z_1}{\sqrt{(z_2/k)}} = \frac{z_1\sqrt{k}}{\sqrt{z_2}}$

is said to follow student's-*t* distribution with *k* degrees of freedom.

The probability curves for the student's-*t* distribution for different degrees of freedom are presented in Fig. 11.4

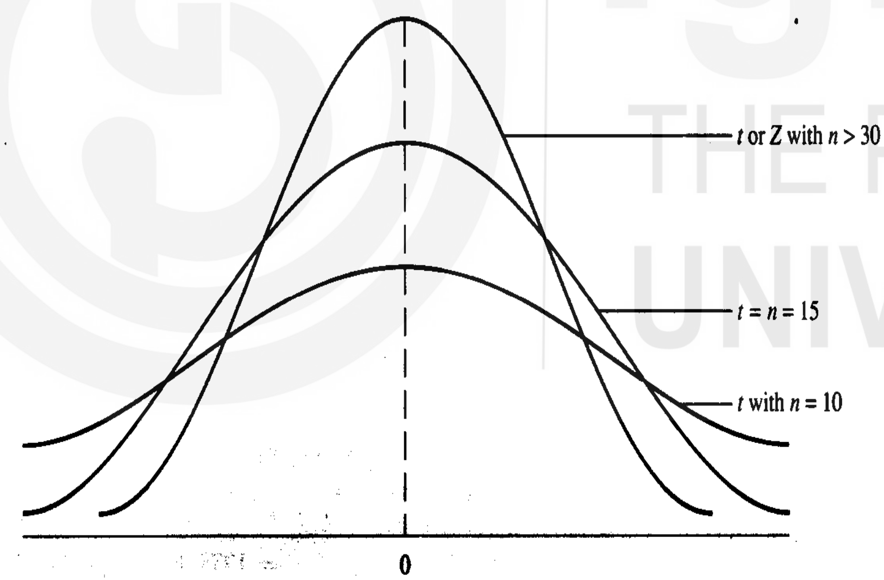

Fig 11.4: Student's-t Probability Curves

We may note the important characteristics of this distribution.

1) As we can see in Fig. 11.4, like the normal distribution, the student's-*t* distribution is also symmetric and its range of variation is also from – ∞ to + ∞; however, it is flatter than the normal distribution. We should also

note here that as the degrees of freedom increase, the student's-t distribution approaches the normal distribution.

2) The mean of the student's-t distribution is zero and its variance is $\frac{k}{(k-2)}$, where, k is the degrees of freedom.

Like the normal distribution, the student's-t distribution is often used in statistical inferences and tests of hypotheses to be discussed in Block-7. The task involves the integration of its density function; which may prove to be tedious. As a result, in this case also, like the normal distribution, a table has been constructed for ready-reference purposes (see Appendix Table A3).

We shall consider an example to see the use of Table A3.

Example 11.5

Given the degrees of freedom equal to 10, what is the probability of obtaining a t value of (i) 2.7638 or greater, (ii) –2.7638 or lower?

Solution:

(i) In Table A3 for student's-t distribution, first, we move down the degrees of freedom column and reach the figure of 10 and then look across the corresponding row and locate the figure of 2.7638. The corresponding lower probability figure of 0.01 is the required probability.

(ii) Since the student's-t distribution is symmetric, the probability of obtaining a t value of –2.7638 or lower is also 0.01.

11.3.4 The *F* Distribution

Another continuous probability distribution that we discuss now is the F distribution.

If z_1 and z_2 are two chi-squared variables that are independently distributed with $k_1$and k_2 degrees of freedom respectively, the variable

$$F = \frac{z_1/k_1}{z_2/k_2}$$

follows s F distribution with k_1 and k_2 degrees of freedom respectively. The variable is denoted by F_{k_{1},k_2} where, the subscripts k_1 and k_2 are the degrees of freedom associated with the chi-squared variables. We may note here that k_1 is called the numerator degrees of the freedom and in the same way, k_2 is called the denominator degrees of freedom.

Some important properties of the F distribution are mentioned below.

1) The F distribution, like the chi-squared distribution, is also skewed to the right. But, as k_1 and k_2 increase, the F distribution approaches the normal distribution.

2) The mean of the F distribution is $k_1/(k_2 - 2)$, which is defined for $k_2 > 2$, and its variance is $\frac{2k_2^2(k_1 + k_2 - 2)}{k_1(k_2 - 2)^2(k_2 - 4)}$ which is defined for $k_2 > 4$.

3) An F distribution with 1 and k as the numerator and denominator degrees of freedom respectively is the square of a student's-t distribution with k degrees of freedom. Symbolically,

$$F_{1,k} = t_k^2$$

4) For fairly large denominator degrees of freedom k_2, the product of the numerator degrees of freedom $k_1$and the F value is approximately equal to the chi-squared value with degrees freedom k_1, i.e., $k_1 F = \chi^2_{k_1}$.

As we have mentioned earlier with reference to other continuous probability distributions, F distribution is also extensively used in statistical inference and testing of hypotheses. Again, such uses also require obtaining areas under the F probability curve and consequently integrating the F density function. However, in this case also our task is facilitated by the provision of the F Table.

11.3.5 Distributions Related to the Normal Distribution

We have already seen from the features of the chi-squared, student's-t and the F distributions that for large degrees of freedom, these distributions approach the normal distribution. Consequently, these distributions are also known as the distributions related to the normal distribution. This relationship between the chi-squared distribution, the student's-t distribution and the F distribution on one hand and the normal distribution on the other has tremendous practical implications. When the degrees of freedom happen to be fairly large, instead of using the chi-squared distribution or the student's-t distribution or the F distribution separately as the situation may demand; we can uniformly apply the normal distribution. As a result, our task gets considerably simplified.

Check Your Progress 2

1) What is the probability of obtaining a χ^2 value of 8 or greater, given the degrees of freedom 20?

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

2) Given the degrees of freedom equal to 25, what is the probability of obtaining a t value of 1.708 or greater?

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

3) Given $k_1 = 10$ and $k_2 = 8$, what is the probability of obtaining an F value of 5.8 or greater?

………………………………………………………………………………………

………………………………………………………………………………………

………………………………………………………………………………………

………………………………………………………………………………………

11.4 LET US SUM UP

In this unit, we studied some continuous probability distributions. Among these distributions, the normal distribution is considered to be the most important one. We have learnt its characteristics and seen its practical applications. We have considered the important concept of a standard normal variate. We have also learnt the technique of using the table for the areas under the standard normal curve for solving problems relating to the normal distribution. Besides the normal distribution, we have considered three other continuous probability distributions. These are: the chi-squared distribution, the student's-t distribution and the F distribution. These three distributions use the notion of the degrees of freedom. So, we have tried to explain the concept of the degrees of freedom. We have learnt the characteristics of these distributions and seen how these distributions can be applied to various situations by using tables relating to these distributions. Finally, we have seen that for fairly large degrees of freedom, these three distributions approach the normal distribution.

11.5 ANSWERS OR HINTS TO CHECK YOUR PROGRESS EXERCISES

Check Your Progress 1

1) a) 0.0552
   b) 0.0668
   c) 0.0062

2) The proportion of persons having heights between 67 inches and 69 inches = 0.1586

   Therefore, the number of persons having heights between 67 inches and 69 inches = 1000 × 0.1586 = 159 (approximately).

3) $$z = \frac{24-34}{5} = -2$$

   From the standard normal table, the area between $z = 0$ and $z = 2$ is 0.4772. Therefore, the area between $z = 0$ and $z = -2$ is 0.4772 (because the standard normal distribution is symmetric). For finding the number of batteries that will have to be replaced, we have to consider the area to the left of $z = -2$, which is equal to the area to the right of $z = 2$. Now, the area to the right of $z = -2$ is $0.5 - 0.4772 = 0.0228$. Therefore, the probability that a battery is defective is 0.0228. Thus out of 5000 batteries, the number of batteries that will have to be replaced $= 0.0228 \times 5000 = 114$.

Check Your Progress 2

1) 0.99
2) 0.05
3) 0.01

ignou
THE PEOPLE'S
UNIVERSITY