UNIT 8: FUNDAMENTALS OF PROBABILITY

UNIT STRUCTURE

8.1 Learning Objectives

8.2 Introduction

8.3 Random Experiment

8.4 Classical Definition of Probability

8.5 Set

- 8.5.1 Subset
- 8.5.2 Sample Space

8.6 Axiomatic Definition of Probability

8.7 Elementary Theorems on Probability

8.8 Let Us Sum Up

8.9 Further Reading

8.10 Answer to Check Your Progress

8.11 Model Questions

8.1 LEARNING OBJECTIVES

After going through this unit, you will be able to:

- explain the concept of random experiment
- discuss the concept of events
- define probability
- explain the concept of sample space
- explain limitations of classical and statistical definition and need for axiomatic definition
- calculate probabilities of events
- define some elementary theorems on probability
- explain the concept of independent events
- evaluate probabilities of independent events.

8.2 INTRODUCTION

The subject probability began in the 17th century through efforts of some mathematician (to quote Fermat and Pascal) to answer questions

concerning games of chance. It was not untill the 20th century that a rigorous mathematical theory based on axioms, definitions and theorems was developed.

All phenomena in nature and society are either deterministic or probabilistic. Till the 18th century only deterministic phenomena were amenable to scientific treatment, but during the last two centuries a great deal of research has been carried out on phenomena involving elements of chance.

As time progresses, theory of probability found its way into many applications, not only in mathematical sciences also in wider fields like agriculture, business, medical science etc. All these applications contributed to further development of the theory.

The term probability or chance we use in our daily talk to indicate the degree of personal belief. We often say “there is no probability of raining today”. The statement just reflects the speaker’s belief without any base on experiment or trial. But in statistics the term is used in different way which is more reliable than the former one.

The theory of probability can be discussed in two different stages. The first one is classical approach. In this approach, probability is defined as the ratio of number of favourable cases when the cases are exhaustive, equally likely and mutually exclusive. The second one is modern approach. It is based on concept of sample space, which is again based on set.

8.3 RANDOM EXPERIMENT

If we perform certain experiment under identical conditions we expect to arrive at results that are essentially same provided we can control the value of the variable that may effect the outcome of the experiment. But in many cases we are not in a position to control the value of some variables. As a result, though under identical conditions experiments are performed, the result will vary from one experiment to the others. These experiments are called random experiments.

Examples of random experiments:

1) Let the experiment be tossing of a coin. The result will be either head (H) or tail (T). But we cannot exactly predict what the result will be. The result will depend on chance. There are various factors that will influence the result and all these factors cannot be controlled. This is an example of random experiment. If coin is tossed twice, the result will be {(H, H), (H, T), (T, H), (T, T)}. This is also example of random experiment.
2) Let us consider the random experiment of throwing a six faced cubical die. As the die is perfect, we are sure that one of the faces will come up with one of the numbers, 1, 2, 3, 4, 5, 6.

Event: An event is a collection of possible outcome of a random experiment. An event may consist of a single outcome or a group of outcomes taken together. If it consists of single outcome then it is called elementary event. An event will be denoted by capital letter of English alphabet.

A series of events $A_1, A_2 \ldots, \ldots, \ldots, A_K$ will be called exhaustive if at least one of them is sure to happen in any trial of random experiment. For example, when a coin is tossed, either head (H) or tail (T) must occur. Hence the event is exhaustive.

Two events A and B are said to be mutually exclusive, if the occurrence of one precludes the occurrence of the other. In other words the two events cannot occur simultaneously. For example in casting a die if 6 occurs then other numbers 1, 2, 3, 4 and 5 cannot occur. Hence it is an example of mutually exclusive events.

A series of events $A_1, A_2 \ldots, \ldots, \ldots, A_K$ are said to be equally likely, if one of them cannot be expected to occur in preference to the others in a single trial of random experiment. For example, in tossing of a coin each face is equally likely to occur.

8.4 CLASSICAL DEFINITION OF PROBABILITY

Consistent with the conditions of a random experiment let ‘n’ be exhaustive, mutually exclusive and equally likely cases and ‘m’ of them are

favourable to an event A, then probability of A is defined by 'm/n' and is denoted by 'P (A)' or 'p' so that–

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

Since $0 \leq m \leq n$, P(A) lies between 0 and 1.

1) When there is no elementary event favourable to A, then P(A) = 0, since m = 0.
2) When all events are favourable to A then P (A) = 1.

Example 1: Let a six faced die be cast. Find the probability that (i) the number shown on the die is odd (ii) number shown on the die is divisible by 2 and 3 (iii) number shown on the die is divisible by 2 and 5.

Solution: When die is cast there are six cases 1, 2, 3, 4, 5 or 6. These are exhaustive, mutually exclusive and equally likely.

i) Of the cases the favourable cases are 1, 3, 5. Let the event be A.

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

ii) In the second case let the event be B then there is only one favourable case viz. 6.

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

iii) In this case let the event be C. Out of the numbers 1, 2, 3, 4, 5, 6 no one is divisible by 2 and 5 simultaneously. Hence no case is favourable to C.

$$P(C) = 0$$

Example 2: A card is drawn at random from a well shuffled pack of cards. Find the probability that it is (i) a king (ii) a queen of spade (iii) a heart.

In a pack of cards there are 52 cards and any card can be drawn and hence total number of cases is 52.

i) Let A denote the first event. There are 4 kings. So number of favourable cases for drawing a king is 4.

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

ii) Let B denote the second event. There are only one queen of spade.

$$P(B) = \frac{1}{52}$$

iii) Let C denote the third event. There are 13 hearts. So favourable cases to C is 13.

$$P(C) = \frac{13}{52} = \frac{1}{4}$$

Limitations of classical definitions: The classical definition requires that n is finite. But there are instances where it may be infinite Secondly we assume elementary events to be equally likely. This is also need not be true always. For example, in tossing of a coin, there are two events head and tail. Unless the coin is unbiased they may not be equally likely. To get a perfectly unbiased coin is a very difficult condition. Hence probability of obtaining a head with any coin cannot be obtained from classical definition.

Statistical definition: To overcome the difficulties encountered in classical definition, a large number of trials of random experiment is considered. Let m be the number of occurrences of an event A associated with n independent trials of random experiment. The probability of the event A is defined by–

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

Example: In tossing of a coin 1000 times head comes up 537 times. The probabiity of head coming up is–

$$\frac{537}{1000} = 0.537$$

This definition of probability also has drawback, because the large number involved in definition is vague.

As both the definitions suffer from certain defects another approach has been put forward to overcome all defects. This approach is known as axiomatic approach of probability.

Before considering this approach we shall introduce a concept, which is as Set. We shall briefly state the necessary results here.

8.5 SET

Set is a concept. It has no definition. Inituitively, a set is any well-defined collection of objects. The objects in a set can be anything. The

numbers, letters, people etc. These objects are called elements or members of the Set.

Examples of Sets:

1) The numbers 19, 23, 29 and 31.
2) The vowels of English alphabet : a, e, i, o and u.
3) The states of North-East India.
4) The girl students Anjana, Binita, Chandana, Deepika and Lina.
5) The district head quarters of Assam.

Sets are normally denoted by capital letters and their elements by small letters.

If A denotes the Set given in example 1 then we write,

$$A = \{19, 23, 29, 31\}$$

Elements of a Set are separated by commas and enclosed in { }. This form of representing a Set is called tabular form. This is also called Roster method.

There is another way of representing a Set. Look at the example 3. Here we have not written the names of the states. Let B denote this Set. Then if x be a state.

$$B = \{x : x \text{ is a state of North-East India}\}$$

This form is Set builder form. This is also called Rule method.

If an object x is a member of a Set A, then we write $x \in A$ [Read as x belongs to A]. If A does not contain x then we write $x \notin A$.

Defining set: A Set is finite if it consists of specific number of different elements, otherwise set is infinite. For example Set of months in a year is finite set. It consists of 12 elements. Set of stars in the sky is infinite it cannot be counted. Set of rivers on the earth is **finite**. Though it is difficult to count still it can be counted. Set of natural number is **infinite Set**.

Universal sets: All Sets under investigation will likely be subsets of a fixed set. This fixed set is universal set and denoted by U.

Equality of sets: Two sets A and B are said to be equal if they have the same members. In other words, every elements which belongs to A also belongs to B and every element which belongs to B also belongs to A, then A and B are equal (A = B).

A set does not change if its elements are repeated.

Let A = {5, 6, 7, 5, 8} and B = {7, 5, 7, 6, 7, 8}. Then A = B.

Null Set: It is mathematically convenient to introduce the concept of null set (Empty Set). A Set which contains no elements is a null set. It is denoted by the symbol ∅.

8.5.1 Subsets

If every element of a Set A is also an element of another Set B, then A is a subset of B. Symbolically it is denoted by A ⊂ B.

If x ∈ A ⇒ x ∈ B, then A ⊂ B.

Let A = {a, b, c} and B = {a, b}

Then B ⊂ A and A will be a super set of B.

From definitions of subsets and equality of sets we see that if A ⊂ B and B ⊂ A then A = B.

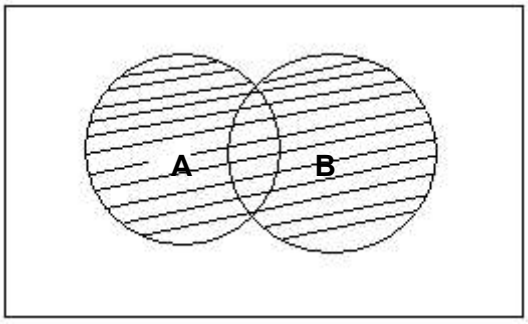
Basic Set operations:

1) **Union of sets:** The union of sets A and B is the set of all elements which belong to A or to B or to both and is denoted by $A \cup B$.

Let A = {1, 2, 3, 4}, B = {2, 4, 6, 8} and C = {1, 3, 5, 7}

Then A ∪ B = {1, 2, 3, 4, 6, 8}, B ∪ C = {1, 2, 3, 4, 5, 6, 7, 8}

In Venn diagram notation:



A ∪ B is the shaded portion.

A ∪ B = {x : x ∈ A or x ∈ B}

A and B are always subsets of A ∪ B.

A ⊂ A ∪ B and B ⊂ A ∪ B



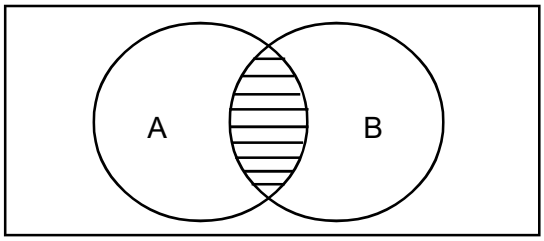
LET US KNOW

Venn diagrams or set diagrams were invented by British mathematician John Venn in 1880. These diagrams show all possible logical relationship between and among different sets that share something in common by using circle.

2) **Intersection of Sets :** The intersection of sets A and B is the set of elements common to A and B and is denoted by $A \cap B$. In the previous example A = {1, 2, 3, 4}, B = (2, 4, 6, 8} and C = {1, 3, 5, 7}.

$A \cap B = \{2, 4\}$, $A \cap C = \{1, 3\}$ and $B \cap C = \varnothing$

In Venn diagram notation,



The shaded portion is $A \cap B$.

$A \cap B$ is subsets of both A and B.

$A \cap B \subset A$ and $A \cap B \subset B$

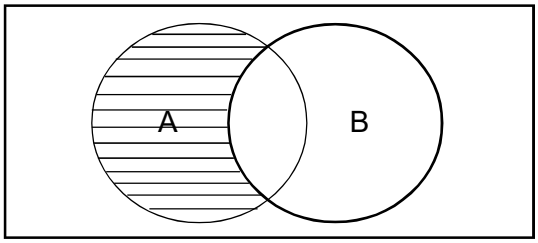
$A \cap B = \{x : x \in A \text{ and } x \in B\}$

If A and B have no common elements then $A \cap B = \varnothing$

In that case A and B are said to be disjoint.

3) **Difference of sets:** The difference of sets A and B is the set of elements which belong to A but not to B and is denoted by A – B.

In the previous example, A – B = {1, 3}



In Venn diagram notation,

Shaded portion is A – B.

Similar definition for B – A

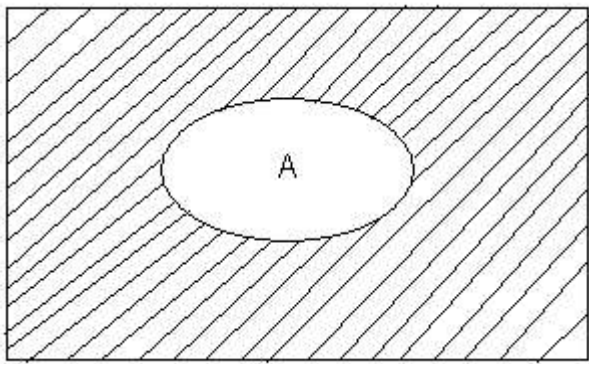
$A - B = \{x : x \in A \text{ and } x \notin B\}$

$A - B \subset A$

4) Complement of a Set: The complement of a Set A is the set of elements which do not belong to A, ie difference of universal set U and A and denoted by A^C.

$$A^C = U - A$$

In Venn diagram notation:



Shaded portion is A^C.

With this background of set we are in a position to define which is known as sample space, which plays significant role in modern development of theory of probability.

Illustrative examples on sets :

1) Let us consider the sets

   $A = \{1, 2, 3, 4\}$ and $B = \{3, 1, 4, 2\}$

   Each elements 1, 2, 3, 4 of A belongs to B

   $$A \subset B$$

   Again each elements 3, 1, 4, 2 of B belongs to A

   $$B \subset A$$

   Hence A = B

2) $A = \{x : 2x = 6\}$ and b=3. Does b belongs to A?

   $A = \{x : 2x = 6\} \Rightarrow A = \{b\}$ when x = 3

   So, $b \in A$

3) Write true and false of the following statements

   i) $8 \in \{1, 3, 5, 7, 9\}$

   ii) If $A = \{1, 4, 6\}$ and $B = \{1, 8\}$ here $A \cap B = \{1\}$

iii) $A \cup A = A$

iv) Set of even numbers is finite

v) The set of letters in the word "follow" and the set of letters in the word "wolf" are different.

Answers: (i) False, (ii) True, (iii) True, (iv) False, (v) False

4) Find union, intersection and diference of the sets

A = {1, b, c, d} and B = {f, b, d, g}

Solution: $A \cup B = \{a, b, c, d, f, g\}$, $A \cap B = \{b, d\}$,

$A - B = \{a, c\}$

5) Let U = {1, 2, 3 ..., ..., ..., 8, 9} and A = {2, 4, 6, 8}

Solution: $A^c = \{1, 3, 5, 7, 9\}$



CHECK YOUR PROGRESS

Q.1: A = {1, 2, 3, 4} and B = {1, 4, 9, 16}.

Fill up the following by $\in$ or $\notin$.

i) 3 ..... A ii) 5 ..... B

iii) 8 ..... B iv) 4 ..... B

Q.2: Which of the following sets are equal : {1, 2, 3}, {3, 2, 1, 3}, (2, 3, 2, 1} and {2, 1, 2}

Q.3: Find all the subsets of {0, 1, 2} and {x, y, z}

Q.4: Find union, intersection and difference of A and B.

Where A = {1, 2, 3, 4, 5} and B = {2, 5, 7}

Q.5: Write True or False of the following:

i) Months of the year in a finite set.

ii) A null set is the set {0}

iii) x = {a, b}, Y = {a, b, c} and Z = {a, b, d}. Then

a) $Z \subset X$ b) $X = Z$ c) $Y \subset X$

iv) A = {1, 0}, then $\{0) \subset A$

v) A = {1, 4, 6} and B = {1, 8} then

a) $A \cup B = \{1, 4, 8\}$ b) $A \cap B = \{1\}$

c) $A - B = \{4, 6, 8\}$

8.5.2 Sample Space

A set S that consists of all possible outcome of a random experiment is called a sample space. Each outcome is a sample point.

Example: Let us consider the result of casting a six-faced die. The outcome will be the appearance of one and only one of the numbers 1, 2, 3, 4, 5, 6, on the upper most face of the die. In other words the six possibilities are such that no two or more of them can occur simultaneously and at least one of them must occur. The space S is {1, 2, 3, 4, 5, 6}. If A denote the event "occurrences of even number" then A = {2, 4, 6} which is a subset of S.

Let us generalise the concept.

Let all possible outcomes of some particular experiment be e_1, e_2, ..., ..., ..., e_n which are such that no two or more of them can occur simultaneously and exactly one of them must occur when the experiment is performed. Any set associated with an experiment which satisfies the above mentioned two properties is called a sample space. The elements or points of the sample space associated with an experiment are called elementary events of the experiment.

If a sample space has finite number of points then it is a finite sample space. If it has many points as there are natural numbers then it is a countably infinite sample space. If it has as many points as there are in $0 \leq x \leq 1$ then it is an infinite sample space.

Examples:

1) If a coin is tossed twice, the four possible results are {HH, HT, TH, TT} which is the sample space. The event that only one head comes up will be {HT, TH} which is a subset of in sample space.
2) From an urn containing 4 balls of different colours red (R), Blue (B), White (W) and Green (G) draw two balls simultaneously. The sample space for this experiment is {RB, RW, RG, BW, BG, WG}.

Note: 1) RB and BR represent the same outcome because we draw balls simultaneously.

Instead of balls drawing simultaneously, let us now draw balls in successions with replacement. The sample space will be {RR, RB, RW, RG, BR, BB, BW, BG, WR, WB, WW, WG, GR, GB, GW, GG}.

2) In this case RB and BR are not same.

8.6 AXIOMATIC DEFINITION OF PROBABILITY

Let S be a sample space. To each event A in a class C of events, a real number P(A) is associated.

P (A) is called probability of the event A if the following axioms are satisfied.

Axiom 1. For any event A in the class C, $P(A) \geq 0$.

Axiom 2. For certain event S in class C, P(S) = 1.

Axiom 3. For any number of mutually exclusive events A_1, A_2, ..., ..., ..., A_n in the class C,

$P(A_1 \cup A_2 \cup \ldots \ldots \ldots) = P(A_1) + P(A_2) + \ldots \ldots \ldots$

In particular for two events, A_1 and A_2

$$P(A_1 \cup A_2) = P(A_1) + P(A_2)$$

Let S be a sample space $\{e_1, e_2, \ldots, \ldots, \ldots, e_n\}$ and $P(\{e_i\}) = p_i$,

i = 1, 2, ..., ..., ..., n such that $0 \leq P_1 \leq 1$ and $\sum_{i=1}^{n} P_i = 1$

Now $P(\{e_i\}) = \frac{1}{n}$

If E be any event consisting of $r(1 \leq r \leq n)$ elementary events then

$P(E) = \frac{1}{n} + \frac{1}{n} + \ldots \ldots \ldots$ r times

$= \frac{r}{n}$

$P(E) = \frac{n(E)}{n(S)}$, where n(E) = number of events in E and

n(S) = number of events in S.

8.7 ELEMENTARY THEOREMS ON PROBABILITY

Some elementary theorems on probability :

Let $S = \{e_1, e_2, ..., ..., ..., e_n\}$ be a sample space of an experiment.

Theorem 1: $P(S) = 1$.

Proof: By definition, $P(S) = P(\{e_1\}) + P(\{e_2\}) + ... + P(\{e_n\})$

$= p_1 + p_2 + ... + p_n$

$= 1$

This theorem is very trivial.

Theorem 2: If A and B are two events such that $A \subseteq B$ then $P(A) \leq P(B)$

Proof: $A \subseteq B \Rightarrow$ B contains all the elementary events of A and possibly a few more.

So, the probability of B is equal to the probability of A plus the sum of the probabilities attached to those elementary events which belongs to B but not to A i.e. to B – A. Since the probability is a non-negative number,

$$P(A) \leq P(B)$$

Theorem 3: For any event A, $0 \leq P(A) \leq 1$

By axiom 1, $P(A) \geq 0$ ... ... (i)

For any event A, $A \subseteq S$

By theorem 2, $P(A) \leq P(S) = 1$ (by theorem 1) ... ... (ii)

$\therefore$ From (i) and (ii), $0 \leq P(A) \leq 1$

Theorem 4: Probability of impossible event is zero.

Let $\varnothing$ denote the impossible event

We know, $S = \cup \varnothing$

$\Rightarrow P(S) = P(S \cup \varnothing) = P(S) + P(\varnothing)$ (by axiom 2)

$\Rightarrow P(\varnothing) = 0$

Let us examine the converse of theorem 1 and theorem 4.

For this let us consider the sample space.

$$S = \{HH, HT, TH, TT\}$$

Let us consider, $P(TH) = \frac{1}{2} = P(HH)$

$$P(HT) = 0\ P(TT)$$

If $A = \{HT, TT\}$ then $P(A) = 0$, but A is not impossible event.

Hence converse of theorem 4 is not true.

Let B = {TH, HH} then P(B) = 1, but B is not sure event.

Hence converse of theorem 1 is not true.

Theorem 5: $P(A_2 - A_1) = P(A_2) - P(A_1)$

We know, $A_2 = A_1 \cup (A_2 - A_1)$, where A_1 and A_2 are mutually exclusive.

$$P(A_2) = P[A_1 \cup (A_2 - A_1)]$$
$$= P(A_1) + P(A_2 - A_1)$$
$$\Rightarrow P(A_2 - A_1) = P(A_2) - P(A_1)$$

Theorem 6: If A^C is the complement of A then $P(A^C) = 1 - P(A)$

We know, $A \cup A^C = S$ and $A \cap A^C = \varnothing$

$A \cap A^C = \varnothing \Rightarrow$ A and A^C are mutually exclusive

Hence, $A \cup A^c = S \Rightarrow P(A \cup A^c) = P(S)$

$$\Rightarrow P(A) + P(A^c) = 1$$
$$\Rightarrow P(A^C) = 1 - P(A)$$

Independent Events: If the probability of B occuring is not affected by the occurence or non-occurence of A then A and B are said to be independent events.

In this case $P(A \cap B) = P(A).P(B)$

Conversely of $P(A \cap B) = P(A).P(B)$ then A and B are independent.

Calculation of probabilities: All these probability theories are used in business to evaluate financial and decision making risk. Probability is used to improve business performnance.

Example 1: In a single cast with two dice find the probability of throwing (i) two aces (ii) doublets (iii) five-six

Solution: Any face of either die may turn up. So sample space consists of 6 × 6 = 36 sample points.

i) Two aces means (6, 6) only one point in sample space.

Probability for the aces = $\frac{1}{36}$

ii) For doublets, the point are (1, 1), (2, 2), (3, 3), (4, 4), (5, 5) and (6, 6), which are six in numbers.

Probability for the doublets = $\frac{6}{36} = \frac{1}{6}$

iii) For five-six, we have two points(5, 6) and (6, 5).

Probability for five-six = $\frac{2}{36} = \frac{1}{18}$

Example 2: Two dice one red and other blue are thrown. Find the probability of the event E where E is–

i) Sum of the number shown by the two dice is 7.

ii) Number on the second dice is greater than the number on the 1st.

Solution: The sample space consists of 6x6=36 elementary events.

i) E_1 = {(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)}

$P(E_1) = \frac{6}{36} = \frac{1}{6}$

ii) E_2 = {(1,2), (1,3), (1,4), (1,5), (1,6), (2,3), (2,4), (2,5), (2,6), (3,4), (3,5), (3,6), (4,5), (4,6), (5,6)}

$P(E_2) = \frac{15}{36} = \frac{5}{12}$

Example 3: A coin is tossed three times in succession. Find the probability of the event that have

i) two or more heads

ii) number of heads equal to the number of tail

iii) exactly one head and two tails.

Solution: The sample space is {(HHH), (HHT), (HTH), (THH), (HTT), (THT), (TTH), (TTT)}

i) E_1 = {(HHH), (HHT), (HTH), (THH)}

$P(E_1) = \frac{4}{8} = \frac{1}{2}$

ii) E_2 = {(HHH), (TTT)}

$P(E_2) = \frac{2}{8} = \frac{1}{4}$

iii) E_3 = {(HTT), (THT), (TTH)}

$P(E_3) = \frac{3}{8}$

Example 4: From a pack of usual playing cards, a card is drawn at random and is noted. Calculate the probabilities of the following events:

i) the drawn card is either a spade or a club.

ii) the drawn card is a picture card (ace is include).

iii) the drawn card is of denomination less than 10 and greater than5.

Solution: The sample space consists of 52 elementary events corresponding to each of the 52 cards that can be drawn from the pack.

i) Let, E_1 be the event that the drawn card is a spade or a club. Then E consists of (13 + 13) = 26 elementary events.

$$P(E_1) = \frac{26}{52} = \frac{1}{2}$$

ii) E_2 be the event that drawn card is a picture card. Here aces are taken as picture card. Then E_2 contains 4 × 4 = 16 elementary events.

$$P(E_2) = \frac{16}{52} = \frac{4}{13}$$

iii) E_3 be the event such that denomination lies between 5 and 10. E_3 consists of 4 × 4 = 16 elementary events.

$$P(E_3) = \frac{16}{52} = \frac{4}{13}$$

Example 5: Two drawings, each of 3 balls, are made from a bag containing 5 white and 8 black balls, the balls not being replaced before the second trial. Find the chance that the first drawing will give 3 white and the second 3 black balls.

Solution: Let E_1 be the event of drawing 3 white balls and E_2 be the event of drawing 3 black balls.

The sample space consists of ${}^{13}C_3$ elementary events for E_1.

$$P(E_1) = \frac{{}^{5}C_3}{{}^{13}C_3}$$

$$= \frac{\frac{5.4}{2.1}}{\frac{13.12.11}{3.2.1}}$$

$$= \frac{\frac{5.4}{2.1}}{\frac{3.2.1}{13.12.11}}$$

$$= \frac{5}{143}$$

For E_2 the sample space consists of 10_{C_3} elementary events.

$$P(E_2) = \frac{8_{C_3}}{10_{C_3}}$$

$$= \frac{\frac{8.7.6}{3.2.1}}{\frac{10.9.8}{3.2.1}}$$

$$= \frac{8.7.6}{3.2.1} \times \frac{3.2.1}{10.9.8}$$

$$= \frac{7}{15}$$

Since E_1 and E_2 are independent,

$$P(E_1E_2) = P(E_1).P(E_2)$$

$$= \frac{5}{143} . \frac{7}{15}$$

$$= \frac{7}{429}$$

Example 6: Two drawings, each of 3 balls, are made from a bag containing 5 white and 8 black balls, the balls being replaced before the second trial find the chance that the first drawing will give 3 white, and the second 3 black balls.

Solution: Let, E_1 be the event of drawing 3 white balls and E_2 be the event of drawing 3 black balls.

Since balls are replaced these sample space for both the events will contained same number of elementary events.

E_1 contains 5_{C_3} elementary events and E_2 contains 8_{C_3} elementary events.

$$P(E_1) = \frac{5_{C_3}}{13_{C_3}}$$

$$= \frac{\frac{5.4}{2.1}}{\frac{13.12.11}{3.2.1}}$$

$$= \frac{5}{143}$$

$$P(E_2) = \frac{8_{C_3}}{13_{C_3}}$$

$$= \frac{\frac{8.7.6}{3.2.1}}{\frac{13.12.11}{3.2.1}}$$

$$= \frac{8.7.6}{13.12.11}$$

$$= \frac{28}{143}$$

Since E_1 and E_2 are independent,

$$P(E_1 \cap E_2) = P(E_1).P(E_2)$$

$$= \frac{5}{143}.\frac{28}{143}$$

$$= \frac{140}{20449}$$

Example 7: A bag contains 5 white and 7 black balls; if two balls are drawn what is the chance that one is white and the other black?

Solution: The sample space consists of $12_{C_2} = \frac{12.11}{2.1} = 66$ elementary events.

Let E_1 be the event of drawing first white and second black balls.

$$P(E_1) = \frac{5}{12}.\frac{7}{11} = \frac{35}{132}$$

Let, E_2 be the event of drawing first black and second white balls.

$$P(E_2) = \frac{7}{12}.\frac{5}{11} = \frac{35}{132}$$

E_1 and E_2 are mutually exclusive. Therefore,

$$P(E_1 \cup E_2) = P(E_1) + P(E_2)$$

$$= \frac{35}{132} + \frac{35}{132}$$

$$= \frac{70}{132}$$

$$= \frac{35}{66}$$

Example 8: Four cards are drawn from a pack of 52 cards. What is the chance that one of each suit is drawn ?

The sample space consists of 15_{C_4} elementary events.

Let E be the event of drawing one card from each suit.

$$P(E) = \frac{13_{C_1} \times 13_{C_1} \times 13_{C_1} \times 13_{C_1}}{52_{C_4}}$$

$$= \frac{13 \times 13 \times 13 \times 13}{\frac{52.51.50.49}{4.3.2.1}}$$

$$= \frac{28561}{270725}$$

Example 9: From an ordinary pack of cards a card is drawn at random. What is the probability that it is–

i) 3 of clubs or 6 of diamonds.

ii) any suit except heart.

iii) a jack or a queen.

iv) either a spade, or a heart or not a picture card.

Solution: i) Let, A denote 3 of clubs, and B denote 6 of diamonds.

Now $A \cap B = \varnothing$

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

$$= \frac{1}{52} + \frac{1}{52}$$

$$= \frac{1}{26}$$

ii) If H denote heart that H/ will denote any suit except heart.

$$P(H) = \frac{13}{52} = \frac{1}{4}$$

$P(H/) = 1 - P(H)$

$$= 1 - \frac{1}{4}$$

$$= \frac{3}{4}$$

iii) Let, A denote that the card is jack & B that is queen. Now

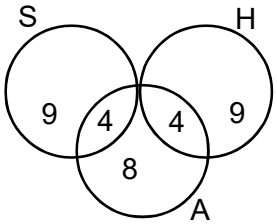
$A \cap B = \varnothing$

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

$$= \frac{4}{13} + \frac{4}{13}$$

$$= \frac{2}{13}$$

iv) Let, S denotes spade, H denotes heart and A denotes picture card. S and H are mutually exclusive but S and A, H and A are not mutually exclusive.



$P(S) = \frac{13}{52}$, $P(H) = \frac{13}{52}$ and $P(A) = \frac{16}{52}$

$P(A') = \frac{36}{52}$, $P(S \cap A') = \frac{9}{52}$

$P(H \cap A') = \frac{9}{52}$

$P(H \cap A') - P(S \cap H) + P(S \cap H \cap A')$

$$P(S \cup H \cup A) = \frac{13}{52} + \frac{13}{52} + \frac{36}{52} - \frac{9}{52} - \frac{9}{52}$$

$[\because P(S \cap H) = 0, P(S \cap H \cap A') = 0]$

$$= \frac{62}{52} - \frac{18}{52}$$

$$= \frac{44}{52}$$

Example 10: Through certain election 3 persons are to be chosen out of five candidates A, B, C, D, E with equal chance being elected. Find the probability that

i) A and B are elected or B and C are elected.

ii) E is elected or C and D are elected.

Solution: Let E_1 be the event that A and B are elected and E_2 be the events that B and C are elected.

S = {ABC, ABD, ABE, ACD, ACE, ADE, BCD, BCE, BDE, CDE}

i) E_1 = {ABC, ABD, ABE}

E_2 = {ABC, BCD, BCE}

$E_1 \cap E_2$ = {ABC}

$P(E_1) = \frac{3}{10}$, $P(E_2) = \frac{3}{10}$, $P(E_1 \cap E_2) = \frac{1}{10}$

$\therefore P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$

$P(E_1 \ E_2) = \frac{3}{10} + \frac{3}{10} - \frac{1}{10} = \frac{5}{10} = \frac{1}{2}$

Let E_3 be the event that E is elected and E_4 be the event that B and C are elected.

ii) E_3 = {ABE, ACE, ADE, BCE, BDE, CDE}

E_4 = {ACD, BCD, CDE}

$E_3 \cap E_4$ = {CDE}

$P(E_3) = \frac{6}{10}$, $P(E_4) = \frac{3}{10}$, $P(E_3 \cap E_4) = \frac{1}{10}$

$\therefore P(E_3EE_4) = P(E_3) + P(E_4) - P(E_3CE_4)$

$P(E_3 \ E_4) = \frac{6}{10} + \frac{3}{10} - \frac{1}{10} = \frac{8}{10} = \frac{4}{5}$

Odds for and odds against: Let E be an event. Sometimes we say that the odds for E are a : b. By this we mean

$$P(E) = \frac{a}{a+b}$$

Thus given the odds for an event, the probability of is event can be calculated. Conversely given the probability of an event, the odds for the event can determined. For example

if P(E) = .7 then, $P(E) = \frac{7}{10} = \frac{7}{7+3}$

Hence odds for E are 7 : 3.

If odds for E is a : b then odds against E is b : a.

Composite Event: When an event, comprises of more than one simple event then the event is called a composite event.

The event $A \cup B$ comprises of the sample points in the whole region bounded by A and B. $A \cup B$ is also denoted by A + B. Similarly the event $A \cap B$ is denoted by AB.

Example 11: From a bag containing 4 white and 5 black balls a man draws 3 balls at random; what are the odds against these being all black?

Solution: The total number of ways in which 3 balls can be drawn is ${}^{9}C_{3}$, and the number of ways of drawing 3 black balls is ${}^{5}C_{3}$; therefore the chance of drawing 3 black balls

$$= \frac{{}^{5}C_{3}}{{}^{9}C_{3}} = \frac{5.4.3}{9.8.7} = \frac{5}{42}$$

Thus the odds against the event are 37 to 5.

Example 12: A party of ten take their seats at a round table. What are the odds against two specified persons (A, B) sitting together?

Solution: A having taken his place, B has a choice of 9 places, 2 of which are next to A, hence the odds against B sitting next to A are as 7 : 2.



CHECK YOUR PROGRESS

Q.6: In a single throw with two dice, find the chances of throwing (i) five, (ii) six.

.....

Q.7: From a pack of 52 cards two are drawn at random; find the chance that one is a king and the other a queen.

.....

Q.8: A bag contains 5 white, 7 black and 4 red balls; find the chance that three balls drawn at random are all white.

.....

Q.9: If four coins are tossed, find the chance that, there should be two heads and two tails.

.....

Q.10: One of two events must happen : given that the chance of the one is two-thirds that of the other, find the odds in favour of the other.

.....



ACTIVITY 8.1

You may perform some random experiments like tossing of coin, casting of a die, drawing cards or some other practical field you come across and calculate probabilities. You may verify the accuracy of the result by repeating the experiment.

.....

.....



8.8 LET US SUM UP

In this unit we have discussed the concept of rondom experiment. In certain experiments, value of the variables cannot be controlled. Therefore, the results of these experiments are not the same. These are known as random experiments. Then we have discussed the concept of probability. Another important topic that we have discussed is the set. Set is a well defined collection of objects which may be numbers, letters, people, animal etc.

Ultimately we have given axiomatic definition. We have discussed the concept of probability and their computational procedure. The discussion is extended to composite events. All are discussed through rules and numerical computation.



8.9 FURTHER READING

1) Gupta, S. K. (1968). *Elements of Probability*. NCERT.
2) Medhi, J. (2005). *Statistical Method*. New Age International Publishers.
3) Mukherjee, Kalyan Kumar (1993). *Probability and Statistics*. New Central Book Agency.
4) Murray, R., Spiegel, Jhon, Schiller, R., Srinivasan, Alu (2005). *Probability and Statistics*. Schaum's Outline Series.



8.10 ANSWERS TO CHECK YOUR PROGRESS

Ans. to Q. No. 1: i) $\in$, ii) $\notin$, iii) $\notin$, iv) $\in$

Ans. to Q. No. 2: First three sets are equal.

Ans. to Q. No. 3: {0}, {1}, {2}, {0,1}, {0,2}, {1,2}, {0,1,2}, $\varnothing$

{x}, {y}, {z}, {x,y}, {x, z}, [y,z], {x, y, z}, $\varnothing$

Ans. to Q. No. 4: $A \cup B = \{1, 2, 3, 4, 5, 7\}$, $A \cap B = \{2, 5\}$,

$A - B = \{1, 3, 4\}$ $B - A = \{7\}$

Ans. to Q. No. 5: i) True, ii) False, iii) (a) False, (b) False, (c) False,

iv) True, v) (a) False, (b) True, (c) False

Ans. to Q. No. 6: $\frac{1}{9}$

Ans. to Q. No. 7: $\frac{5}{36}$

Ans. to Q. No. 8: $\frac{1}{56}$

Ans. to Q. No. 9: $\frac{3}{8}$

Ans. to Q. No. 10: 2 to 3



8.11 MODEL QUESTIONS

Q.1: There are three events A, B, C one of which must, and only one can, happen; the odds are 8 to 3 against A, 5 to 2 against B; find the odds against C.

Q.2: If two balls are drawn from a bag containing 2 white, 4 red and 5 black balls, what is the chance that

i) they are both red?

ii) one is red and the other black?

Q.3: Two coins are tossed simultaneously. Find the probability of getting–

i) exactly two heads or head is the first coin or both,

ii) getting head in first and load is the 2nd coin or both.

Q.4: A card is drawn at random from a pack of playing cards. Find the probability that it is

i) a ten or a spade.

ii) neither a 4 nor a club.

iii) either a black card or an ace or both.

iv) either a spade or a picture card.

Q.5: Two dice are thrown simultaneously. Find the probability that the sum is

i) divisible by either 2 or 5 but not by both.

ii) does not exceed 6.

iii) greater than 10.

iv) divisible by 2 and 5 both.

Q.6: If A and B are any events prove that

$P(A \cap B) \leq P(A) \leq P(A \cup B) \leq P(A) + P(B)$

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

Q.7: If $P(A) = \frac{1}{2}$, $P(B) = \frac{1}{4}$, $P(A \cup B) = \frac{3}{8}$, find P(AB).

Q.8: There are three events A, B, C one of which must and only one can happen. The odds are 8 to 3 against A, 5 to 2 against B. Find the odds against C.

Q.9: A experiment consists of drawing 3 cards in succession from a pack of cards. A_1 is the event king on first draw, A_2 king on the 2nd and A_3 king is the 3rd. State the meaning of

i) $P(A_1 \cap A^{/})$ ii) $P(A_1 \cup A_2)$ iii) $P(A_1^{/} \cup A_2^{/})$

iv) $P(A^{/} \cap A_2^{/} \cap A_3^{/})$ v) $P[(A_1 \cap A_2) \cup (A_2^{/} \cap A_3)]$

*** ***** ***

UNIT 9: CONDITIONAL PROBABILITY

UNIT STRUCTURE

9.1 Learning Objectives

9.2 Introduction

9.3 Conditional Probability

9.4 Multiplication Theorem on Probability

9.5 Total Probability

9.6 Baye's Theorem

9.7 Let Us Sum Up

9.8 Further Reading

9.9 Answers to Check Your Progress

9.10 Model Questions

9.1 LEARNING OBJECTIVES

After going through this unit, you will be able to:

- define conditional probability
- know multiplication theorem of probability
- define law of probability
- know Baye's theorem.

9.2 INTRODUCTION

In unit 7, we discussed about probbility and some important theorems related to probability. In this unit, we will discuss the concept of conditional probability. We will also discuss dependent and independent events. Multiplication theorems for dependent and independent events will also be discussed in this unit. Finally, we will disuss the law of probability and baye's theorem.

9.3 CONDITIONAL PROBABILITY

In unit 7, we have discussed the methods of finding the probability of events. Supose, we have two events from the same sample space in a

random experiment, does the occurrence of one of the events affect the probability of the other event? Let us see this situation with the following examples

1) Consider a random experiment of throwing a die.

Let S be the sample space and A be the event of of getting an odd numbers.

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

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

Suppose we consider another event B such that B be the event of getting the numbers greater than 3. Then B = {4, 5, 6}.

When we say that B has already occurred,the event of getting 'numbers is less than or equal to 3' is not possible. Now, the number of elements of of the sample space is reduced to 3 i.e. {4, 5, 6}. Out of these three points, two points are even number {4, 6}.

Thus, given that if the event B has already occurred, then $P(A) = \frac{2}{3}$.

Hence, the probability of A given that B has already occurred is denoted by $P\left(\frac{A}{B}\right)$.

2) Consider the experiment of tossing three fair coins. The sample space of the experiment is S = {HHH, HHT, HTH, THH, HTT, THT, TTH, TTT}.

Let E be the event 'at least two tails appear' and F be the event 'first coin shows head'.

Now, suppose we are given that the first coin shows head, i.e. F occurs, then we have to find the probability of occurrence of E. With the information of occurrence of F, we are sure that the cases in which first coin does not result into a head should not be considered while finding the probability of E. This information reduces our sample space from the set S to its subset F for the event E. In other words, the additional information really amounts to telling us that the situation may be considered as being that of a new random experiment for

which the sample space consists of all those outcomes only which are favourable to the occurrence of the event F. Now, the sample point of F which is favourable to event E is HTT.

Thus, Probability of E considering F as the sample space = $\frac{1}{4}$.

Probability of E given that the event F has occurred = $\frac{1}{4}$.

This probability of the event E is called the *conditional probability of E given that F has already occurred*, and is denoted by P (E|F).

3) Consider the random experiment of drawing a card from a deck of 52 cards. Suppose we draw a queen in successive draws. In first draw the probability of drawing a queen is $\frac{4}{52}$. If the card is not replaced before the second draw, then the event of getting queen in the second draw depends on the first event. Now, there will be only 3 queens. The probability of getting a queen in second draw will also be $\frac{3}{52}$.

In the above examples,we observe that the occurance of one event depends on another event which has already happend. Hence,an event E is said to be a dependent event if the occurnce of A depends on another event F which has already happend. It is denoted by P(E|F).

The above examples leads to the definition of conditional probability as:

Let E and F be two events defined on a sample space S.

Let P(F) > 0, then the conditional probability of E given that F has already occurred denoted by P(E|F) and mathematically written as:

$$P(E|F) = \frac{P(E \cap F)}{P(F)}, \quad P(F) > 0$$

Similarly, $P(F|E) = \frac{P(E \cap F)}{P(E)}$, $P(E) > 0$

Note: 1) P(E|F) and P(F|E) can also be expressed as follows:

$P(E|F) = \frac{n(E \cap F)}{n(F)}$ and $P(F|E) = \frac{n(E \cap F)}{n(E)}$.

2) P(E|F) is read as probability of event E given event F.

Laws of Conditional probability:

a) If E_1 and E_2 be any two events of a sample space S and A be any event of S such that $P(A) \neq 0$, then

$P(E_1 \cup E_2|A) = P(E_1|A) + P(E_2|A) - P(E_1 \cup E_2|A)$

We have, $P(E_1 \ R_2|A)$

$$= \frac{P[(E_1 \cup E_2) \cap A]}{P(A)}$$

$$= \frac{P(E_1 \cap A) \cup P(E_2 \cap A)}{P(A)}$$

$$= \frac{P(E_1 \cap A) + P(E_2 \cap A) - P(E_1 \cap E_2 \cap A)}{P(A)}$$

$$= \frac{P(E_1 \cap A)}{P(A)} + \frac{P(E_2 \cap A)}{P(A)} - \frac{P(E_1 \cap E_2 \cap A)}{P(A)}$$

$= P(E_1|A) + P(E_2|A) - P(E_1 \cap E_2|A)$

b) $P(E'|F) = 1 - P(E|F)$

We have, $P(S|F) = 1$

$\Rightarrow P(E' \cup E|F) = 1$

$\Rightarrow P(E'|F) + P(E|F) = 1$

$\Rightarrow P(E'|F) = 1 = 1 - P(E|F)$

Independent Events: Events are said to be independent if the occurrence of one does not affect the others. In the experiment of tossing a fair coin, the occurrence of the event 'head' in the first toss is independent of the occurrence of the event ' head' in the second toss and third toss.

Example 1: A card is chosen at random from an ordinary deck of 52 cards.

Let E be the event that the cards is an ace and F the event that the card is a club. Then E and F are independent.

Dependent Events: If the occurrence of one event influences the occurrence of the other, then the second event is said to be dependent on the first. In the above example, if we do not replace the first ball drawn, this

will change the composition of balls in the bag while making the second draw and therefore the event of 'drawing a red ball' in the second will depend on event (first ball is red or white) occurring in first draw. Similarly, if a person draw a card from a full pack and does not replace it, the result of the draw made afterwards will be dependent on the first draw.

Example 2: Two fair dice are tossed. Let E be the event that the first die is a 3, F the event that the sum is 6 and G the event that sum is 7. Then E and F are dependent, but E and G are independent.

Note: If the events A and B are independent, that is the probability of occurrence of any one of them P(E|F) = P(E) and P(F|E) = P(F).

9.4 MULTIPLICATION THEOREM ON PROBABILITIES

Now, we discuss multiplication theorem on probabilities for both independent and dependent events.

Multiplication theorem on probabilities for independent events: If two events E and F are independent, the probability that both of them occur is equal to the product of their individual probabilities.

i.e. $P(E \cap F) = P(E).P(F)$.

Note: 1) The theorem can be extended to three or more independent events. If E, F, G, …, …, ... be independent events, then $P(E \cap F \cap G) = P(E)P(F)P(G)$.

2) If E and F are independent then the complements of E and F are also independent. i.e. $P(\overline{E} \cap \overline{F}) = P(\overline{E})P(\overline{F})$.

Multiplication theorem for dependent events: If E and F be two dependent events, i.e. the occurrence of one event is affected by the occurrence of the other event, then the probability that both A and B will occur is $P(E \cap F) = P(E)P(F|E)$.

Note: In the case of three events E, F, G.

$P(E \cap F \cap G) = P(E)P(F|E)P(G|E \cap F)$

i.e., the probability of occurrence of E, F and G is equal to the probability of E times the probability of F given that E has occurred, times the probability of G given that both E and F have occurred.

Ilustrative Examples:

Example 1: Let E and F are two events,such that

$P(E) = 0.8$, $P(F) = 0.6$, $P(E \cap F) = 0.5$.

Find the value of (i) $P(E \cup F)$, (ii) $P(F|E)$, (iii) $P(E|F)$.

Solution: (i) $P(E \cup F) = P(E) + P(F) - P(E \cap F)$

$= 0.8 + 0.6 - 0.5$

$= 0.9$

ii) $P(F|E) = \frac{P(E \cap F)}{P(E)} = \frac{0.5}{0.8} = \frac{5}{8}$

iii) $P(E|F) = \frac{P(E \cap F)}{P(F)} = \frac{0.5}{0.6} = \frac{5}{6}$

Example 2: If $P(E) = 0.5$, $P(F) = 0.3$ and $P(E \cap F) = 0.15$. Find $P(E|F)$.

Solution: By definition of conditional probability

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

$$= \frac{0.15}{0.3}$$

$$= 0.5$$

Example 3: If $P(E) = 0.9$; $P(F|E) = 0.8$, find $P(E \cap F)$.

Solution: By definition of conditional probability

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

$\Rightarrow P(F \cap E) = P(E).P(F|E)$

$\Rightarrow 0.9 \times 0.8 = 0.72$

Example 4: Find the chance of drawing 2 white balls in succession from a bag containing 5 red and 7 white balls, the balls drawn not being replaced.

Solution: Let E be the event that ball drawn is white in the first draw and F be the event that ball drawn is white in the second draw.

$\therefore P(F \cap E) = P(E).P(F|E)$

Here $P(E) = \frac{7}{12}$, $P(F|E) = \frac{6}{11}$

$$\therefore P(F \cap E) = \frac{7}{12}.\frac{6}{11} = \frac{7}{22}$$

Example 5: In an examination 30% of the students have failed in Mathematics, 20% of the students have failed in chemistry and 10% have failed in both Mathematics and Chemistry. A student is selected at random (i) What is the probability that the students has failed in Mathematics if it is known that he has failed in Chemistry?

Solution: Let E be the event of failing in Mathematics.

$\therefore P(E) = 30\% = 0.3$

Let F be the event of failing in Chemistry

$\therefore P(F) = 20\% = 0.2$

Let $E \cap F$ be the event of failing in both subjects.

$\therefore P(E \cap F) = 10\% = 0.1$

The probability that the student has failed in Mathematics given that he has failed in Chemistry is:

$$P(E|F) = \frac{P(E \cap F)}{P(F)} = \frac{0.1}{0.2} = 0.5$$

Example 6: A bag contains 5 white and 3 black balls.Two balls are drawn at random one after the other without replacement.Find the probability that both balls are black.

Solution: Probability of drawing a black ball in the first attempt is:

$$P(E) = \frac{3}{5+3} = \frac{3}{8}$$

Probability of drawing the second black ball given that the first ball drawn is black:

$$P(F|E) = \frac{2}{5+2} = \frac{2}{7}$$

$\therefore$ The probability that both balls drawn are black is given by

$$P(E \cap F) = P(A) \times P(F|E) = \frac{3}{8} \times \frac{2}{7} = \frac{3}{28}$$



CHECK YOUR PROGRESS

Q.1: Two fair coins are tossed. Given that at least one of them is heads, what is the probability that both of them are heads?

Q.2: Suppose that a box contains 8 red balls and 4 white balls. We randomly draw 3 balls from the box without replacement. Find the probability that all 3 are red.

Q.3: A die is rolled, If the outcome is an odd number. What is the probability that it is prime?

Q.4: A family has two children. Find the probability that both children are girls given that at least one of them is a girl.

9.5 TOTAL PROBABILITY

Theorem of total probability: If E_1 and E_2 are mutually exclusive and exhaustive events and $P(E_1) \neq 0$, $P(E_2) \neq 0$, then for any event A of S,

$$P(A) = P(E_1)P(A|E_1) + P(E_2)P(A|E_2)$$

Proof: Since E_1 and E_2 are mutually exclusive and exhaustive events, we have $E_1 \cup E_2 = S$ and $E_1 \cup E_2 = \varnothing$.

$\therefore A = A \cap S$

$= A \cap (E_1 \cup E_2)$

$= (A \cap E_1) \cup (A \cap E_2)$ ... ... (1)

Now, $(A \cap E_1) \cup (A \cap E_2) = A \cap (E_1 \cup E_2)$

$= A \cap \varnothing = \varnothing$

$\therefore A \cap E_1$ and $A \cap E_2$ are mutually exclusive and exhaustive events.

$\therefore$ From (1), we have

$P(A) = P(A \cap E_1) + P(A \cap E_2)$

$= P(E_1)P(A|E_1) + P(E_2)P(A|E_2)$

[multiplication theorem of probability]

If E_1, E_2, E_3 are mutually exclusive and exhaustive events and $P(E_1) \neq 0$, $P(E_2) \neq 0$, $P(E_3) \neq 0$, then for any event A of S,

$$P(A) = P(E_1)P(A|E_1) + P(E_2)P(A|E_2) + P(E_3)P(A|E_3)$$

9.6 BAYE'S THEOREM

The concept of conditional probability discussed earlier takes into account information about the occurrence of one event to predict the probability of another event. This concept can be extended to revise probabilities based on new information and to determine the probability that a particular effect was due to specific cause. The procedure for revising these probabilities is known as Bayes theorem. The Principle was given by Thomas Bayes in 1763. By this principle, assuming certain prior probabilities, the posteriori probabilities are obtained.

Bayes' Theorem or Rule:

Statement: Let $E_1, E_2, E_3, ..., ..., ..., E_n$ be a set of n mutually exclusive and collectively exhaustive events and $P(E_1), P(E_2), P(E_3), ,,,, ..., ..., P(E_n)$, are their corresponding probabilities. If A is another event such that P(A) is not zero and the priori probabilities $P(A|E_1)$, i = 1, 2, ..., ..., ..., n are also known. Then

$$P(E_i|A) = \frac{P(E_i)P(A \mid E_i)}{P(E_1)P(A \mid E_1) + P(E_2)P(A \mid E_2) + ... + P(E_n)P(A \mid E_n)}$$

$$\text{i.e., } P(E_i|A) = \frac{P(E_iP(A \mid E_i)}{\sum_{i=1}^{n} P(E_i)P(A \mid E_i)}$$

Proof: Let S be the sample space of the random experiment.

The events $P(E_1), P(E_2), P(E_3), ..., ..., ..., P(E_n)$ being exhaustive.

$\therefore S = E_1 \cup E_2 \cup ... \cup E_n$

$A = A \cap S \quad [A \subset S]$

$= A \cap (E_1 \cap E_2 \cap ... \cap E_n)$

$= (A \cap E_1) \cup (A \cap E_2) \cup ... (A \cap E_n)$ [Distributive law]

$\Rightarrow P(A) = P(A \cap E_1) + P(A \cap E_2) + ... + ... + ... + P(A \cap E_n)$

$\Rightarrow P(E_1)P(A|E_1) + P(E_2)P(A|E_2) + ... + ... + ... + P(E_n)P(A|E_n)$

$$= \sum_{i=i}^{n} P(E_1)P(A \mid E_i) \qquad \text{... ... ... (1)}$$

Now, $P(A \cap E_i) = P(A)P(E_i|A)$

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

$$= \frac{P(E_i)P(A \cap E_i)}{\sum_{i=1}^{n} P(E_i)P(A \mid E_i)} \quad \text{[using (1)]}$$

Note: 1) The probabilities $P(E_1)$, $P(E_2)$, $P(E_3)$, ,,,, ..., ..., $P(E_n)$ are called as the 'a prori probabilites'.

2) The probabilities $P(E_1|A)$, $P(E_2|A)$, ..., ..., ..., $P(E_n|A)$ are called'posteriori probabilities'.

Example 1: A purse contains 6 silver coins and 3 gold coins.Another purse contains 4 silver coins and 5 gold coins.A purse is selected at random and a coin is drawn from it.What is the probability that it is a silver coin.

Solution: Let Let E = Event of selecting first purse

F = Event of selecting the second purse

A = Event of getting a silver coin

$\therefore P(E) = \frac{1}{2}$ and $P(F) = \frac{1}{2}$

Also, $P(E|A) = \frac{6}{9} = \frac{2}{3}$ and $P(F|A) = \frac{4}{9}$

$\therefore$ Required probability $P(A) = P(E)P(A|E) + P(F)P(A|F)$

$$= \frac{1}{2} \times \frac{6}{9} + \frac{1}{2} \times \frac{4}{9}$$

$$= \frac{10}{18}$$

$$= \frac{5}{9}$$

Example 2: A box contains two coins. One coin is heads on both sides and the other is heads on one side and tails on the other. One coin is selected from the box at random and the face of one side is observed. If the face is heads, what is the probability that the other side is heads ?

Solution: Let E = Event of selecting two headed coin

F = Event of selecting the other coin

(one head and one tail)

A = Event of getting the second side also head.

Now, we have to find P(E|A)

Here, $P(E) = P(F) = \frac{1}{2}$

$P(A|E) = 1$, $P(A|F) = \frac{1}{2}$

By Baye's theorem, we have

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

$$= \frac{\frac{1}{2}\times 1}{\frac{1}{2}\times 1+\frac{1}{2}\times\frac{1}{2}}$$

$$= \frac{2}{3}$$

Example 3: Assume that a factory has two machines. Past records show that machine I produces 20% of the items and machine II produces 80% of the items. Further, it shows that 6% of the items produced by machine I are defective and only 1% of the items produced by machine II were defective. If a defective item is drawn at random,what is the probability that it was produced by machine I?

Solution: Let E_1, E_2 be the events of drawing an item produced by machines I and II repectively. Let A denote the event of drawing a defective item.

We have $P(E_1) = \frac{20}{100} = \frac{1}{5}$, $P(E_2) = \frac{80}{100} = \frac{4}{5}$

Also, we have $P(A|E_1) = \frac{6}{100}$, $P(A|E_2) = \frac{1}{100}$

From Bayes' theorem, the required probability is

$$P(E_1|A) = \frac{P(E_1).P(A|E_1)}{P(E_1)P(A|E_1)+P(E_2)P(A|E_2)}$$

$$= \frac{\frac{1}{5}.\frac{6}{100}}{\frac{1}{5}.\frac{6}{100}+\frac{4}{5}.\frac{1}{100}} = \frac{6}{6+4} = 0.6$$

Example4: A bolt is manufactured by 3 machines A, B and C. A turns out twice as many items as B and machines B and C produce equal number of items. 2% of bolts produced by A and B are defective and 4% of bolts produced by C are defective. All bolts are put into one stock pile and 1 is chosen from this pile.What is the probability that it is defective?

Solution: Let E_1, E_2, E_3 be the events of bolts manufactured by machine A, B, C repectively.

Here we have to apply total theorem on probability.

B and C produce equal no. of bolts and A produces two times that of B.

$$\therefore\ P(E_1) = 50\% = \frac{50}{100} = 0.50$$

$$P(E_2) = 25\% = \frac{25}{100} = 0.25$$

$$P(E_3) = 25\% = \frac{25}{100} = 0.25$$

Let D be the event of getting defective bolt.

$$P(D|E_1) = \frac{2}{100} = 0.02$$

$$P(D|E_2) = \frac{2}{100} = 0.02$$

$$P(D|E_3) = \frac{4}{100} = 0.04$$

By total theorem on probability.

$$P(D) = P(E_1)P(D|E_1) + P(E_2)P(D|E_2) + P(E_3)P(D|E_3)$$
$$= 0.5 \times 0.02 + 0.25 \times 0.02 + 0.25 \times 0.04$$
$$= 0.0250$$

Example 5: A bag A contains 2 white and 3 red balls and bag B contains 4 white and 5 red balls.One ball is drawn at random at random from one of the bags and is found to be red,find the probability that it is drawn from bag B.

Solution: Here given that the ball drawn is red and we have to find the probability of the ball coming from bag B.

Let E_1 be the event of selecting bag A.

$$\therefore P(E_1) = \frac{1}{2}$$

Let E_2 be the event of selecting bag B.

$$\therefore P(E_2) = \frac{1}{2}$$

Let D be the event of selecting red ball.

Probability of drawing red ball from bag A is:

$$P(D|E_1) = \frac{3}{5}$$

Similarly, probability of drawing red ball from bag B is:

$$P(D|E_2) = \frac{5}{9}$$

We have to find the red ball was drawn from bag B. i.e., $P(E_2|D)$.

By Baye's Theorem:

$$P(E_2|D) = \frac{P(E_2)P(D \mid E_2)}{P(E_1)P(D \mid E_1) + P(E_2)P(D \mid E_2)}$$

$$= \frac{\frac{1}{2}.\frac{5}{9}}{\frac{1}{2}.\frac{3}{5} + \frac{1}{2}.\frac{5}{9}}$$

$$= \frac{\frac{5}{18}}{\frac{3}{10} + \frac{5}{18}}$$

$$= \frac{\frac{5}{18}}{\frac{27+25}{90}}$$

$$= \frac{\frac{5}{18}}{\frac{52}{90}}$$

$$= \frac{25}{52}$$

Example 6: In a bolt factory machines A, B, C manufacture repectively 25%, 35%, 40% of the total. Of their output 5%, 4%, 2% are defective bolts. A bolt is drawn at random and found to be defective. What is the probability that it was manufactured by machine B ?

Solution: Let E_1, E_2, E_3 be the events of bolts manufactured by machine A, B, C repectively.

$$P(E_1) = 25\% = \frac{25}{100} = 0.25$$

$$P(E_2) = 35\% = \frac{35}{100} = 0.35$$

$$P(E_3) = 40\% = \frac{40}{100} = 0.40$$

Let D be the event of drawing a defective bolt.

$$\therefore\ P(D|E_1) = \frac{5}{100} = 0.05 \quad \text{[defective from 'A']}$$

$$P(D|E_2) = \frac{4}{100} = 0.04 \quad \text{[defective from 'B']}$$

$$P(D|E_3) = \frac{2}{100} = 0.02 \quad \text{[defective from 'C']}$$

Now, we have to find the probability of defective bolt manufactured by machine B. i.e., $P(E_2|D)$.

By Baye's Theorem:

$$P(E_2|D) = \frac{P(E_2)P(D\,|\,E_2)}{P(E_1)P(D\,|\,E_1)+P(E_2)P(D\,|\,E_2)+P(E_3)P(D\,|\,E_3)}$$

$$= \frac{0.35\times 0.04}{0.25\times 0.05+0.35\times 0.04+0.4\times 0.02}$$



CHECK YOUR PROGRESS

Q.5: Assume that a factory has two machines. Past records show that machine I produces 20% of the items and machine II produces 80% of the items. Further, it shows that 6% of the items produced by machine I are defective and 1% of the items produced by machine II are

defective. If a defective item is drawn at random, what is the probability that it was produced by machine I?

Q.6: Three machines A, B, C produce repectively 60%, 30%,10% of the total number of items of a factory. The percentage of respective outputs of these machines are repectively 2%, 3% and 4%. An item is selected at random and is found to be defective.Find the probability that the item was produced by machine C.

Q.7: A bag X contains 2 white and 3 red balls and a bag Y contains 4 white and 5 red balls. One ball is drawn at random from one of the bags and is found to be red. Find the probability that it was drawn from bag Y.

Q.8: Urn A contains 3 red and 4 white balls and Urn B has 4 red and 5 white balls. One urn is selected at random and one ball is taken out of it. Thrown ball is red. What is the probability that it was taken from urn B?

Q.9: A factory has three production lines I, II and III contributing 20%, 30% and 50% respectively, to its total output. The percentages of substandard items produced by lines I, II and III are, respectively, 15, 10 and 2. If an item chosen at random from the total output is found to be substandard, what is the probability that the item is from line I?



9.7 LET US SUM UP

- For two events E and F, the probability of **an** event **E** under the condition that F has already occumh is the **conditional probability P(E|F)** of **E** and is given by $P(E|F) = \frac{P(E \cap F)}{P(F)}$.
- Two events for which the occurrence of one has no influence on the occurrence or non-occurrence of the other are **independent** events, otherwise, they are **dependent.**

- If the events E and F are independent and P(E) > 0, P(F) > 0, then $P(E \cap F) = P(E).P(F)$ and $P(E|F) = P(E)$, $P(F|E) = P(F)$.



9.8 FURTHER READING

1) Rao. G. Shanker, *Probability and Statistics*. Universitis Press.
2) R. S. N. Pillai and Bagavathi. *Statistics*. New Delhi: S. Chand and Company Ltd.



9.9 ANSWERS TO CHECK YOUR PROGRESS

Ans. to Q. No. 1: Let E be the event that both of them are heads and F the event that at least one of them is heads.

Then, $P(E) = \frac{3}{4}$, $P(E \cap F) = \frac{1}{4}$, $P(F) = \frac{1}{4}$

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

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

$$= \frac{1}{3}$$

By using the definition of conditional probability, we can easily get that– $P(E \cap F) = P(F).P(E|F) = \frac{1}{4} \times \frac{1}{3} = \frac{1}{12}$

Ans. to Q. No. 2: For i = 1; 2; 3, let E_i be the event that the i-th ball is red.

Then, $P(E_1 \cap E_2 \cap E_3) = P(E_1).P(E_2|E_1).P(E_3|E_1E_2)$

$$= \frac{8}{12}.\frac{7}{11}.\frac{6}{10} = \frac{14}{55}$$

Ans. to Q. No. 3: When a die is rolled, the sample is S = {1, 2, 3, 4, 5, 6}.

Let E = Event of getting an odd number = {1, 3, 5}

F = Event of getting a prime number = {2, 3, 5}

Then, $E \cap F = \{3, 5\}$

$\therefore P(E) = \frac{3}{6} = \frac{1}{2}$, $P(F) = \frac{3}{6} = \frac{1}{2}$ and $P(E \ F) = \frac{1}{3}$

P(getting a prime already when an odd number)

= P(getting a prime number | getting an odd number)

= P(E|F)

= P(E|F)

$= \frac{P(E \cap F)}{P(F)}$

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

$= \frac{2}{3}$

Ans. to Q. No. 4: The sample space S is given by

S = {(b, b), (b, g), (g, b), (g, g)

Let E : Event that both children are girls

F : Event that at least one of the child is a girl.

Then E = {(g, g) and F = {b, g), (g, b), (g, g)}

$E \cap F = \{(g, g)\}$

$\therefore P(E) = \frac{3}{4}$, $P(E \cap F) = \frac{1}{4}$

$\therefore$ Required probability is $P(E|F) = \frac{P(E \cap F)}{P(F)}$

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

$= \frac{1}{3}$

Ans. to Q. No. 5: Let E_1, E_2 be the events of drawing an item produced by machines I and II repectively. Let A denote the event of drawing a defective item.

We have $P(E_1) = \frac{20}{100} = \frac{1}{5}$, $P(E_2) = \frac{80}{100} = \frac{4}{5}$

Also, we have $P(A|E_1) = \frac{6}{100}$, $P(A|E_2) = \frac{1}{100}$

From Bayes's Theorem, the required probability is

$$P(E_1|A) = \frac{P(E_1)P(A \mid E_1)}{P(E_1)P(A \mid E_1) + P(E_2)P(A \mid E_2)}$$

$$= \frac{\frac{1}{5}.\frac{6}{100}}{\frac{1}{5}.\frac{6}{100} + \frac{4}{5}.\frac{1}{100}}$$

$$= \frac{6}{6+4}$$

$$= 0.6$$

Ans. to Q. No 6: Let E, F, G denote the events in which the items are selected by machines A, B, C repectively.

Then, we have $P(E) = \frac{60}{100} = 0.6$, $P(F) = \frac{30}{100} = 0.3$,

$$P(G) = \frac{10}{100} = 0.1$$

Let D denote the event in which the item selected at random is a defective item. Then

$P(D|E) = \frac{2}{100} = 0.02$, $P(D|F) = \frac{3}{100} = 0.03$, $P(D|G) = \frac{4}{100} = 0.04$

Applying Bayes's theorem, we obtain

$$P(G|D) = \frac{P(G).P(D \mid G)}{P(E)P(D \mid G) + P(F)P(D \mid F) + P(G)P(D \mid G)}$$

$$= \frac{(0.1)(0.04)}{(0.6)(0.02) + (0.3)(0.03) + (0.1)(0.04)}$$

$$= \frac{0.004}{0.012 + 0.009 + 0.004}$$

$$= \frac{0.004}{0.025}$$

$$= 0.16$$

Ans. to Q. No. 7: Let E_1 : the ball is drawn from bag X

E_2 : the ball is drawn from bag Y

and A; the ball is red.

Here $P(E_1) = P(E_2) = \frac{1}{2}$

Also, $P(A|E_1)$ = P (a red ball is drawn from bag X) = $\frac{3}{5}$

$P(A|E_2)$ = P (a red ball is drawn from bag Y) = $\frac{5}{9}$

By Baye's Theorem, we have

$$P(E_2|A) = \frac{P(E_1)P(A \mid E_1)}{P(E_1)P(A \mid E_1) + P(E_2)P(A \mid E_2)}$$

$$= \frac{\frac{1}{2} \times \frac{5}{9}}{\frac{1}{2} \times \frac{3}{5} + \frac{1}{2} \times \frac{5}{9}}$$

$$= \frac{25}{52}$$

Ans. to Q. No. 8: Required probability = $\frac{28}{45}$ [By Bayes's Theorem]

Ans. to Q. No. 9: By Bayes formula

$$P(E_1|A) = \frac{P(E_1)P(A \mid E_1)}{P(E_1)P(A \mid E_1) + P(E_2)P(A \mid E_2)}$$

$$= \frac{0.2 \times 0.15}{0.2 \times 0.15 + 0.3 \times 0.10 + 0.5 \times 0.02}$$

= 0.069



9.10 MODEL QUESTIONS

Q.1: E and F are two events for which P(E) = $\frac{1}{2}$, P(F) = $\frac{1}{3}$, and P(E $\cap$ F) = $\frac{1}{4}$. Find (i) P(E|F), (ii) P(E $\cup$ F) and (iii) $P(E'|F')$.

Q.2: Three persons A, B, C supply raw mterials to a factory in the proportion $\frac{1}{2}, \frac{1}{3}, \frac{1}{6}$ and 5%, 6%, 8% of the materials supplied by them repectively are found to be defective. A material supplied is selected at random from the total supply. What are the probabilities that material was supplied by A, or by B or by C?

Q.3: One of three identical coins is unbiased while the remaining two are biased with probabilities of getting head in a toss are $\frac{1}{3}$ and $\frac{2}{3}$. One of the coins selected randomly shows head when tossed. Find the probability of the coin to be unbiased.

Q.4: Boys from the districts of Kamrup, Nagaon and Jorhat come to a college in Assam in the proportions 50%, 30% and 20% repectively. 80%, 60% and 40% of the boys comingrepectively from the districts wear Dhooti. A boy selected randomly from all the boys of the college is found to wear Dhooti. What is the probability that he comes kamrup?

Q.5: Three trains X, Y, Z carry repectively 25%, 35%, 40% of the total passengers. Of the passengers carried by the tains 5%, 4% and 3% respectively are ladies. If a pssengers chosen at random out of all the passengers of the trains found to be a lady. Prove that the probability that she was carried by X, or Y or Z are repectively $\frac{25}{69}$, $\frac{28}{69}, \frac{16}{69}$.

Q.6: Three are three groups of children. The first group consists of 3 girls and 1 boy, the second consists of 2 girls and 2 boys and the third consists of 1 girl and 3 boys. One child is picked up at random from each group. Find the probability of selecting 1 girl and 2 boys.

Q.7: A bag contains 5 white and 7 red balls. If a ball is drawn twice at random successively, find the probability that one ball is white and the other is red.

Q.8: In a running competition, the probabilities of the two competitiors of winning the race is $\frac{1}{3}$ and $\frac{1}{5}$. What is the probability that both of them will not win ?

Q.9: One card is drawn from a well shuffled pack of 52 cards.Deduce the probability of drawing.

i) a red card or an ace or both

ii) a diamond or an ace or both.

iii) king of diamond or king of heart or both.

Q.10: 4 men play with a pack of cards containing 52 cards. What is the probability that the 4 kings are held by only one of the players?

*** ***** ***

UNIT 10: RANDOM VARIABLES AND ITS PROBABILITY DISTRIBUTION

UNIT STRUCTURE

10.1 Learning Objective

10.2 Introduction

10.3 Random Variable

- 10.3.1 Definition or Random Variable
- 10.3.2 Discrete Random Variable
- 10.3.3 Continuous Random Variable

10.4 Probability Distribution

- 10.4.1 Probability Mass Function
- 10.4.2 Probability Density Function

10.5 Mathematical Expectation and Varience

- 10.5.1 Mathematical Expectation
- 10.5.2 Properties of Mathematical Expectation
- 10.5.3 Variance of Random Variable
- 10.5.4 Properties of Variances
- 10.5.5 Illustrated Examples

10.6 Let Us Sum Up

10.7 Further Reading

10.8 Answers to Check Your Progress

10.9 Model Questions

10.1 LEARNING OBJECTIVES

After going through this unit, you will be able to:

- distinguish between two types of experiments viz. deterministic experiments and random experiments
- define what is a random variable and discuss when we get a discrete random variable or a continuous random variable
- have knowledge about probability distributions
- discuss the concept ot probability mass function and probability density function along with their properties

- know about mathematical expectations and variance of a random variable which are vital in the study of any distribution, along with their properties and applications
- realise how different measures of location and dispersion as discussed in earlier chapters can be expressed with the help of moments.

10.2 INTRODUCTION

The experiments that are performed can be classified into two broad categories. In one type of experiment the outcomes of the experiments remain same whenever and wherever the experiment is performed. In the second type of experiment, the outcomes are found to vary every time the experiment is performed.

The first types of experiments are known as deterministic experiments, whereas the second types are known as random experiments. As we have already discussed, in this type of experiment, the experimenter knows the set of all possible outcomes, but cannot say with certainity which outcome will turn out at a particular time. If we want to denote thses various outcomes of a random experiment with the help of a variable, then it becomes a random variable.

In this unit we shall learn about random variables, the different types of random variables and their mathematical expectation and some related theorems. We shall also study about moments and try to understand the difference between raw moments, i.e. the moment about any arbitary point and central moment, the moment about mean.

10.3 RANDOM VARIABLE

Most of the random experiments we encounter have outcomes that can be interpreted in terms of real numbers, such as marks obtained by a student in an examination, number of accidents in a day etc. On the other hand, sometimes the outcomes of an experiment can take any value in an interval of real numbers. For example, the heights of students in a class

can take any value within a specified range. Sometimes, the outcomes of a random experiment may be qualitative, e.g. tossing of a coin may result in head or in tail. We may denote head by '1' and tail by '0' and in this way the outcomes of a random experiment, quantitaive or qualitative can be expressed by a real number. these numerical values of the outcomes that changes from experiment to experiment are called random variables. So we can formally define a random variable as follows:

10.3.1 Definition of Random Variable

The real numbers which are associated with the outcomes of a random experiment are called random variables.

It can also be defined as a variable which takes a definite set of values with a probability associated with each of its values is called a random variable. That is why it is also called a chance variable or a stochartic variable.

Random variables may be of two types:

i) Discrete random variables

ii) Continuous random variables

Note: If X and Y are two random variables, then X + Y, X – Y are also random variables. Again, if, a and b are constants, then ax + y, x + by, ax + by are also random variables.

10.3.2 Discrete Random Variable

If a random variable x assumes only a finite number or countably infinite number of values, then it is called a discrete random variable. So a random variable x can take finite number of values x_1, x_2, ..., ..., ..., x_n or it can take countably infinite number of values x_1, x_2, ..., ..., ... The number of letters received by a post office during a particular time period, the number of vehicles arriving at a toll bridge, the number of machines breaking down on a given day etc are some examples of discrete random variables.

10.3.3 Continuous Random Variable

If a random variable is such that it takes any value within a given interval, then it is called a continuous random variable. in this case we say that x takes any value in an interval, say. (a, b), i.e. $a \leq x \leq b$; i.e. x is a continuous random variable in the interval (a, b). Amount of rainfall in a rainy season, heights of individuals, the time between arrival of customers at a servic system etc. some examples of continuous random varible.

10.4 PROBABILITY DISTRIBUTION

The distribution obtained by taking all possible values of a random variable along with their respective probabilities is called a probability distribution. When the random variable is discrete we get discrete probability distribution and when the variable under consideration is continuous, we get a continuous probability distribution. The discrete probability distribution is called probability mass function and the continuous probability distribution is called probability density function.

10.4.1 Probability Mass Function

Let x denote a discrete random variable which takes the values $x_1, x_2, ..., ..., ..., x_n$ with respective probabilities $p_1, p_2, ..., ..., ..., p_n$. Let, $p_i = P(x = x_i) = p(x_i)$, i = 1, 2, ..., ..., ..., n be the probability that x takes the value xi. Then p(xi) is called the Probability Mass Function (p.m.f.) of x if it satisfies the following conditions.

i) $p(x_i) \geq 0$ for all i = 1, 2, ..., ..., ..., n

ii) $\sum_{i=1}^{n} p(x_i) = 1$

In case, the discrete random variable x is countably infinite, then, $p(x_i) = P(x = x_i)$, i = 1, 2, ..., ..., ..., n is said to be the p.m.f. of x if

i) $p(x_i) \geq 0$ for all i = 1, 2, ..., ..., ...

ii) $\sum_{i=1}^{\infty} p(x_i) = 1$

Illustration: Let x denote the uppermost face while throwing a dice. Then x can take the values 1, 2, 3, 4, 5 or 6, each with probability $\frac{1}{6}$. This can be written as,

$p(x_i) = P(x = x_i) = \frac{1}{6}$, i = 1, 2, ..., ..., ..., 6

Here, $p(x_i) > 0$

and $\sum_{i=1}^{6} p(x_i) = 1$

Thus $p(x_i)$ satisfies both the conditions for a p.m.f.

10.4.2 Probability Density Function

The probability distribution of a continuous random variable x is defined by f(x) is called the probability density function (p.d.f.). Let, x be a continuous random variable in the interval (a,b), then the function f(x) = P(X = x) is said to be the p.d.f. of x, if it satisfies the following conditions.

i) $f(x) \geq 0 \ \forall$ x in (a, b)

ii) $\int_a^b f(x)dx = 1$

Illustration: Consider the function, $f(x) = \begin{cases} a, 0 \leq x \leq 5 \\ 0, \text{otherwise} \end{cases}$

For f(x) to be a p.d.f., the condition $\int_a^b f(x)dx = 1$ must be satisfied, which is true if,

$$\int_0^5 a dx = 1$$

$$\Rightarrow a = \frac{1}{5}$$

Since a > 0, the function $f(x) \geq 0$. Thus f(x) satisfies both the conditions for a p.d.f.

10.5 MATHEMATICAL EXPECTATION AND VARIANCE

10.5.1 Mathematical Expectation

The mathematical expectation of a discrete random variable is the sum of the product of the values taken by the random variable and their respective probabilities. Let us consider the discrete random variable x, which assumes the values $x_1, x_2, ..., ..., ..., x_n$ with respective probabilities $p_1, p_2, ..., ..., ..., p_n$, then the mathematical expectation of the random variable x is denoted by E(x) and is given by,

$$E(x) = \sum_{i=1}^{n} x_i p_i$$

Provided the series is convergent and $\sum_{i=1}^{n} p_i = 1$

If x is a continuous random variable with p.d.f. f(x), $-\infty < x < \infty$, then the mathematical expectation of the random variable x is given by,

$$E(x) = \int_{-\infty}^{\infty} xf(x)dx \text{ , provided } \int_{-\infty}^{\infty} f(x)dx = 1$$

Note: The expectation of a random variable x represents the mean of the probability distribution of x.

Mathematical Expectation of a function of a random variable:

If g(x) is a function of a random variable x, then the expectation of g(x) is given by,

$$E[g(x)] = \sum_{i=1}^{n} g(x_i)p(x_i) = \sum_{i=1}^{n} g(x_i)p_i$$

where, x is a discrete random variable taking values $x_1, x_2, ..., ..., ..., x_n$ with respective probabilities $p_1, p_2, ..., ..., ..., p_n$.

Similarly, when x is a continuous random variable in the interval $(-\infty, \infty)$, with p.d.f. f(x), $-\infty < x < \infty$, then,

$$E[g(x)] = \int_{-\infty}^{\infty} g(x)f(x)dx$$

10.5.2 Properties of Mathematical Expectation

The following are the important properties of an expected value of a random variable:

1) The expected value of a constant is the constant itself.
   i.e. E(c) = c, for every constant c.
2) The expected value of the product of a constant a and a random variable x is equal to the times the expected value of the random variable, i.e. E(ax) = a.E(x).
3) The expected value of a linear function of a random variable is same as the linear function of its expectation,
   i.e. E(a + bx) = a + b.E(x).
4) The expected value of the sum of the two random variables is equal to the sum of their individual expected values.
   i.e. E(x + y) = E(x) + E(y).
5) The expected value of the product of two independent random variables is equal to the product of their individual expected values. i.e. E(xy) = E(x).E(y).

10.5.3 Variance of a Random Variable

Variance is an important characteristic of a random variable. It provides the measure of dispersion of the random variable. The variance of a random variable x is denoted by V(x) or σ^2 and is given by,

$$\sigma^2 = v(x) = E\{(x) - E(x)\}^2$$
$$= E(x^2) - \{E(x)\}^2$$

Note: The positive square root of variance is the standard deviation.

10.5.4 Properties of Variance

1) The variance of a constant is zero. i.e. V(c) = 0.
2) The variance of the product of a constant and a random variable is equal to the variance multiplied by the square of the constant.
   i.e. $V(cx) = c^2.V(x)$.

3) The variance of the sum or difference of two independent random variables. i.e. $V(x \pm y) = V(x) + V(y)$.

10.5.5 Illustrated Examples

1) A random variable x has the following probability distribution

x:	–2	–1	0	1	2	3
p(x):	0.1	0.1	0.2	0.2	0.3	0.1

Calculate E(x) and V(x)

Solution:

$E(x) = \sum_x xp(x)$

$= -2 \times 0.1 + (-1) \times 0.1 + 0 \times 0.2 + 1 \times 0.2 + 2 \times 0.3 + 3 \times 0.1$

$= 0.8$

$V(x) = E(x^2) - \{E(x)\}^2$

Now, $E(x^2) = \sum_x x^2 p(x)$

$= (-2)^2 \times 0.1 + (-1)^2 \times 0.1 + 0 + 1^2 \times 0.2 + 2^2 \times 0.3 + 3^2 \times 0.1$

$= 2.8$

$\therefore V(x) = 2.8 - (-0.8)^2$

$= 2.8 - 0.64$

$= 2.16$

2) In an office there are 10 HCL and 20 HP computers. One computer is selected at random. Find the probability distribution of selecting an HP computer. What is the expected value of selecting a HP computer.

Solution: Let x be a random variable such that,

x = 1, if a HP computer is selcted

= 0, if a HP computer is not selected.

$\therefore$ the probability distribution of x is,

x:	0	1
P(X = x):	$\frac{10}{30}$	$\frac{20}{30}$
	$= \frac{1}{3}$	$= \frac{2}{3}$

$\therefore$ E (an HP computer is selected) $= 0 \times \frac{1}{3} + 1 \times \frac{2}{3}$

$$= \frac{2}{3}$$

3) Eighty percent patients of a hospital are suffering from hypertension. One patient is randomly selected. Find the probability distribution of patients who are suffering from hypertension and find mean of the distribution.

Solution: Let x be a random variable such that,

x = 1, if the patient suffers from hypertension

= 0, otherwise

Hence the probability distribution of x is

x:	0	1
p(X = x):	0.2	0.8

and $E(x) = 0 \times 0.2 + 1 \times 0.8$

$= 0.8$

4) The monthly demand for transistors have the following probability distribution

Demand (n):	1	2	3	4	5	6
Probability (p):	0.10	0.15	0.20	0.25	0.18	0.12

Determine the expected demand for transistors. If the cost of producing n transistors is given by C = 10,000 + 500n, determine the expected cost.

Solution: Expected demand for transistor = E(n)

$= \sum np$

$= 1 \times 0.10 + 2 \times 0.15 + 3 \times 0.20 + 4 \times 0.25 + 5 \times 0.18 + 6 \times 0.12$

$= 3.62$

$E(C) = E(10,000 + 500n)$

$= 10,000 + 500\ E(n)$

$= 10,000 + 500 \times 3.62$

$= 11,810$

5) Proprieter of a food stall has invented a new item of food delicacy. He has calculated that the cost per piece will be Rs. 1, but he

wants to sell it for Rs. 3 per piece. It is, however, perishable and the goods unsold at the end of the day is a total loss. If demand is a variable with the following frequency distribution, calculate his net profit if he makes 15 pieces.

No. of pieces demanded:	10	11	12	13	14	15
Probability:	0.07	0.10	0.23	0.38	0.12	0.10

Solution: Units produced = 15

Total cost price = 15 x 1 = Rs. 15

sell price peer unit = Rs. 3

Unit demanded	Profit
10	Rs. (10×3) – 15 = Rs. 15
11	Rs. (11×3) – 15 = Rs. 18
12	Rs. (12×3) – 15 = Rs. 21
13	Rs. (13×3) – 15 = Rs. 24
14	Rs. (14×3) – 15 = Rs. 27
15	Rs. (15×3) – 15 = Rs. 30

$\therefore$ Expected net profit = Σ profit x probability

= 15 × 0.07 + 18 × 0.10 + 21 × 0.23 + 24 × 0.38 + 27 × 0.12 + 30 × 0.10

= Rs. 3.45



CHECK YOUR PROGRESS

Q.1: Check true or false:

i) The probability distribution of a random variable x is as follows

X = x:	1	2	3	
P(X = x):	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{4}$	(T/F)

ii) If c is a constant then, (a) E(c) = c (T/F), (b) V(c) = c (T/F)

iii) If x and y are two independent random variables, then,

a) V(x + y) = V(x) + V(y) (T/F)

b) V(x – y) = V(x) – V(y) (T/F)

iii) If x and y are two independent random variables, then, E(xy) = E(x).E(y)

iv) A continuous random variable can take any value within a given range.

Q.2: Define random variable. What are the important properties of a random variable.



10.6 LET US SUM UP

- The experiments can be classified into two broad categories - deterministic experiment and random experiment.
- The real numbers which are associated with the outcomes of a random experiment are called random variables.
- Random variables may be of two types - discrete random variable and continuous random variable.
- The distribution obtained by taking all possible values of a random variable along with their respective probabilities is called a probability distribution.
- The probability distribution for discrete random variable is called probability mass function (p.m.f.) and the same for continuous random variable is known as probability density function (p.d.f.).
- The mathematical expectation of a discrete random variable is the sum of the product of the values taken by the random variable and their respective probabilities.
- The expectation of the random variable x represents the mean of the probability distribution of x.
- The variance of a random variable provides a measure of dispersion of the random variable.
- The expectation of a constant quantity is the constant itself, whereas variance of a constant is zero.
- The expectation of the sum of two or more random variables is the sum of the individual random variables.

- The expectation of the product of two independent random variables is the product of their expectations.
- The variance of the addition or subtraction iof two independent random variables is equal to the sum of their variances.



10.7 FURTHER READING

1) Feller, W. (1968). *An Introduction to Probability Theory and its Application*. Vol. I. Wiley Eastern.
2) Goon, A.M., Gupta, M. K. and Dasgupts, B. (1997). *An Ouline of Statistical Theory*. Vol. I. World Press.
3) Gupta, S. C., Kapoor, V. K. (2009). *Fundamentals of Mathematical Statistics*. Sultan Chand & Sons.
4) Mukhopadhyay, P. (1996). *Mathematical Statistics*. New Central Book Agency (P) Ltd.
5) Mukhopadhyay, P. (1991). *Theory of Probability*. New Central Book Agency (P) Ltd.



10.8 ANSWERS TO CHECK YOUR PROGRESS

Ans. to Q. No. 1: i) False (Total probability exceeds 1); ii) (a) True, (b) False; iii) (a) True, (b) False; iv) True; v) True

Ans. to Q. No. 2: The real numbers whcih are associated with the outcomes of a random experiment are called random variables.

The properties of random variable are as follows:

i) If X is a random variable there, $\frac{1}{x}$, x^2 are also random variables.

ii) If x is a random variable and a and b are constants then ax+b is also a random variable.

iii) If x and y are two random variables, then x + y and x – y are also random variables.



10.9 MODEL QUESTIONS

Q.1: Fill in the blanks:

i) Heights or weights of individuals can be represented by a ..... random variable.

ii) If x and y are two random variables then ax+by is a ....., where a and b are constants.

iii) If x and y are independent, then E(xy) = ...

iv) $V(x \pm y) = V(x) + V(y)$, when x and y are ..... variables.

v) V(ax + b) = ..., where a and b are constants.

vi) E(a) = ... and V(a) = ..., where a is a constant.

Q.2: Can you find the mathematical expectation of a random variable x, if x takes the values 1, 2, 3 with corresponding probabilities $\frac{1}{4}, \frac{1}{2}$ and $\frac{3}{4}$.

Q.3: Discuss discrete and continous random variables with examples.

Q.4: What do you mean by mathematical expectation of a function of a random variable. State addition and multiplication law of mathematical expectation.

Q.5: A random variable x has the following probability distribution,

x:	–2	–1	0	1	2	3
f(x):	0.1	2k	0.2	2k	0.3	2k

Find the value of k.

Q.6: Write some important properties of mathematical expectation.

*** ***** ***

UNIT 11: THEORETICAL PROBABILITY DISTRIBUTIONS (DISCRETE VARIABLE)

UNIT STRUCTURE

11.1 Learning Objectives

11.2 Introduction

11.3 Moment Generating Function

- 11.3.1 Definition
- 11.3.2 Generation of Moments
- 11.3.3 Properties of Moment Generating Function
- 11.3.4 Illustrative Examples

11.4 Binomial Distribution

- 11.4.1 Derivation of Binomial Distribution
- 11.4.2 Definition
- 11.4.3 Moments of Binomial Distribution
- 11.4.4 Moment Generating Function of Binomial Distribution
- 11.4.5 Fitting of Binomial Distribution
- 11.4.6 Properties of Binomial Distribution
- 11.4.7 Importance of Binomial Distribution
- 11.4.8 Illustrative Examples

11.5 Poisson Process

- 11.5.1 Poisson Distribution as a Limiting Case of Binomial Distribution
- 11.5.2 Moments of Poisson Distribution
- 11.5.3 Moment Generating Function
- 11.5.4 Fitting of Poisson Distribution
- 11.5.5 Properties of Poisson Distribution
- 11.5.6 Application of Poisson Distribution
- 11.5.7 Illustrative Examples

11.6 Let Us Sum Up

11.7 Further Reading

11.8 Answers To Check Your Progress

11.9 Model Questions

11.1 LEARNING OBJECTIVES

After going through this unit, you will able to:

- obtain moment generating function for any given distribution
- generate moments from any probability distribution
- learn some important properties of moment generating function
- derive binomial distribution from a coin tossing experiment
- derive poisson distribution as a limiting form of binomial distribution
- compute moments of binomial and poisson distribution
- fit binomial and poisson distribution to real life date
- understand the application of the concept of these probability distributions to real life data.

11.2 INTRODUCTION

So far we have brought out the nature of discrete and continuous probability distributions in general and also have defined their mean and variances. In this present unit we shall discuss about how the moments of a particular distribution as discussed in chapter 6, can be generated by a special kind of function called the moment generating function.

Again we know that the rule generating the probability distribution is the same whether we want to find the probability of the number of heads when three coins are lossed simultaueously or for the number of red balls in case of an urn containing 3 red and 2 black balls, out of which, say, two balls are drawn at random with replacement. Given this basic characteristic of a discrete distribution, our main concern is to examine the outcoms generated by an experiment and choose a distribution that best describes those outcomes.

Under this consideration we shall discuss some discrete probability distributions applicable to many real life situations in the subsequent sections of this unit. The distributions we are going to discuss here are binomial and poisson distributions.

11.3 MOMENT GENERATING FUNCTION

The moments of a probability distribution play an important role in theoretical and in applied statistics. With the help of moments we can measure the central tendency of a set of observations, their variability, their asymmetry and the height of the peak their curves would make. That is the knowledge of moments enable us to determine the probability distribution completely. So it would be very useful if a function could be found which will help in generating moments. One such function is moment generating function. So we can define a moment generating function as follows:

11.3.1 Definition

A function that is used to generate moments of a random variable X is called the moment generating function, denoted by Mx(t) and is given as:

$$M_x(t) = E(e^{tx})$$

When X is a discrete random variable with p.m.f. p(x)

$$M_x(t) = E(e^{tx}) = \sum_x e^{tx} p(x)$$

When x is a continous random variable in the interval (a,b) with p.d.f. f(x),

$$M_x(t) = E(e^{tx}) = \int_a^b e^{tx} f(x)dx$$

11.3.2 Generation of Moments

We have,

$M_x(t) = E(e^{tx})$

$$= E\left(1 + tx + \frac{t^2x^2}{2!} + \frac{t^3x^3}{3!} + ... + \frac{t^rx^r}{r!} + ...\right)$$

$$= E(1) + tE(x) + \frac{t^2}{2!}E(x^2) + \frac{t^3}{3!}E(x^3) + ... + \frac{t^r}{r!}E(x^r) + ...$$

$\therefore \mu_1'$ = co-efficient of t in $M_x(t)$

$\therefore \mu_2'$ = co-efficient of $\frac{t^2}{2!}$ in $M_x(t)$

$\therefore \mu_3'$ = co-efficient of $\frac{t^3}{3!}$ in $M_x(t)$

Thus in general we see that,

$\therefore \mu_r'$ = co-efficient of $\frac{t^r}{r!}$ in $M_x(t)$

Another convenient way of obtaining moments is to differentiate $M_x(t)$ sucussively and then put t = 0.

So if we differentiate $M_x(t)$, one time, with respect to t and then put t = 0, we get μ_1',

i.e. $\mu_1' = \left[\frac{d}{dt}M_x(t)\right]_{t=0}$

Differenting $M_x(t)$ times with respect to t and then put t = 0

we get μ_α'.

$\mu_2' = \left[\frac{d^2}{dt^2}M_\lambda(t)\right]_{t=0}$

So in general, we get the raw moment of order r, μ_r' as:

$\mu_r' = \left[\frac{d^r}{dt^r}M_r(t)\right]_{t=0}$

11.3.3 Properties of Moment Generating Function

i) If X is a random variable and C is a constant, there,

$M_{C_x}(t) = M_\lambda(ct)$

Proof: L.H.S. = $M_{C_x}(t) = E\left(e^{tcx}\right)$

$= E\left(e^{ctx}\right)$

$= M_x(ct)$

= R.H.S.

ii) Additive property of moment generating function.

The moment generating function of sum of independent random variables is equal to the product of their moment generating

functions. In symbol, if $X_1, X_2, ..., ..., ..., X_n$ are independent random variables, then,

$Mx_1 + x_2 + ... + x_n(t) = Mx_1(t).Mx_2(t). ... .Mx_n(t)$

Proof: L.H.S. $= Mx_1 + x_2 + ... + x_n(t)$

$= E\left[e^{t(x_1+x_2+...+x_n)}\right]$

$= E\left[e^{tx_1+tx_2+...+tx_n}\right]$

$= E\left(e^{tx_1}\right)E\left(e^{tx_2}\right)....E\left(e^{tx_n}\right)$

$\because$ $X, X_2, ..., ..., ..., X_n$ are independent

$= Mx_1(t).Mx_2(t). ... .Mx_n(t)$

= R.H.S.

iii) Effect of change of origion and scale on m.g.f.

$$M_{\frac{x-a}{h}}(t) = e^{\frac{-at}{h}} M_x\left(\frac{t}{h}\right)$$

Proof: $M_{\frac{x-a}{h}}(t) = E\left[e^{t\left(\frac{x-a}{h}\right)}\right]$

$= E\left[e^{\frac{tx}{h}} e^{\frac{-at}{h}}\right]$

$= e^{\frac{-at}{h}} E\left[e^{\frac{t}{h}(x)}\right]$

$= e^{\frac{-at}{h}} M_x\left(\frac{t}{h}\right)$

= R.H.S.

Note: i) The moment generating function doesnot always exist.

ii) In some cares the moment generating function exists, but it cannot generate moments.

11.3.4 Illustrative Example

1) Let X be a random variable with p.m.f. $p(x) = q^x p$, x = 0, 1, 2, ... find the moment generating function and have find its mean and variance.

Solution: $M_x(t) = E(e^{tx})$

$$= \sum_{x=0}^{\alpha} e^{tx}.q^x p$$

$$= p\sum_{x=0}^{\alpha} \left(qe^t\right)^x$$

$$= p[1 + qe^t + (qe^t)^2 + ...]$$

$$= p(1 - qe^t)^{-1}$$

$$\text{Mean} = \mu_1' = \left[\frac{d}{dt}M_x(t)\right]_{t=0}$$

$$= [(-1)p(1-qe^t)^{-2}(-qe^t)]_{t=0}$$

$$= pq(1-q)^{-2}$$

$$= \frac{pq}{p^2}$$

$$= \frac{p}{p}$$

$$\mu_\alpha' = \left[\frac{d^2}{dt^2}M_x(t)\right]_{t=0}$$

$$= \left[\frac{d}{dt}\left\{pqe^t\left(1-ge^t\right)^{-2}\right\}\right]_{t=0}$$

$$= pq[et(1-ge^t)^{-2} + e^t(-2)(1-ge^t)^{-3}(-ge^t)]_{t=0}$$

$$= pq[(1-q)^{-2} + 2q(1-q)^{-3}]$$

$$= \frac{pq}{p^2} + \frac{2pq^2}{p^3}$$

$$= \frac{q}{p} + \frac{2q^2}{p^2}$$

The expansion used here is called binomial expansion with negative index, which is $(1-x)^{-1} = 1 + x + x^2 + x^3 + ... + ... + ...$

$$\therefore \text{ variance} = \mu_\alpha = \mu_\alpha' - \mu_\alpha'^2 = \frac{q}{p} + \frac{2q^2}{p^2} - \frac{q^2}{p^2}$$

$$= \frac{q}{p} + \frac{q^2}{p^2} = \frac{q(p+q)}{p^2} = \frac{q}{p^2}$$

2) Final moment generating function of a continuous random variable x having the p.d.f. $f(x) = \theta e^{-\theta}$, $x \geq 0$

Hence find its mean and variance.

Solution: $M_x(t) = E(e^{tx})$

$$= \int^{\alpha} e^{tx} \theta e^{-\theta x} dx$$

$$= \theta^0 \int^{\alpha} e^{(t-\theta)x} dx$$

$$= \theta^0 \left[\frac{e(t-\theta)x}{t-\theta} \right] \alpha 0$$

$$= \theta \left[\frac{1}{t-\theta} (0-1) \right]$$

$$= \frac{-\theta}{t-\theta}$$

$$= \frac{\theta}{\theta - t}$$

$$= \left(1 - \frac{t}{\theta}\right)^{-1}$$

$$= 1 + \frac{t}{\theta} + \left(\frac{t}{\theta}\right)^2 + \left(\frac{t}{\theta}\right)^3 + \ldots + \ldots + \ldots$$

$\therefore$ μ_1' = mean co-efficient of t in $M_x(t) = \frac{1}{\theta}$

μ_2' = mean co-efficient of t in $M_x(t) = \frac{2}{\theta^2}$

$\therefore$ variance = $\mu_2 = \mu_2' - \mu_1'^2$

$$= \frac{2}{\theta^2} - \left(\frac{1}{\theta}\right)^2$$

$$= \frac{1}{\theta^2}$$



CHECK YOUR PROGRESS

Q.1: State true or false:

i) Since $M_x(t)$ generates moments, it is called moment generating function. (T/F)

ii) If X an Y are independent r. V S then
$M_{x+y}(t) = M_x(t) + M_y(t)$ (T/F)

iii) Moment generating function always exits. (T/F)

iv) A moment generating function, if exists, determences the distribution function. (T/F).

11.4 BINOMIAL DISTRIBUTION

The binomial distribution, as derived by James Bernoulli in the year 1700, deals with population whose elements or outcomes can be devided into two catigories with reference to presence or absence of a particular attribute or characteristic. For example, the answer to a question may be correct or incorrect, results in an examination may be pass or fail, the factory workers may be educated or uneducated.

That is the outcome in each of these situations may be viewed in terms of presence or absence of an attribute. That is we consider only the two outcomes one possessing and the other not possissing the attribute. A convenient method is to denote them as ‘success’ and ‘failure’ according as the presence or absence of the attribute.

The mathematical background for binomial distribution describes descrete date resulting from an experiment called. Bernoulli trials, which are repeted finite number of times. Let us now device binomial distribution from a finite number of Bernoulli treals.

Note: A random experiment whoss outcomes has been classified into two categories, called ‘success’ and ‘failure’ is called Bernoulli trial. For example, tossing of a coin may result in a head or in a tail etc.

11.4.1 Derivation of Binomial Distribution

Consider a coin tossing experiment which is repeated, say, n number of times. When we get a head we call it a success and when we get a tail, we call it a failure, success is denoted by ‘s’ and failure by ‘f’. Probability of success is p and that of failure is q.

$$P(s) = p,\ P(f) = q,\ q = 1 - p.$$

Suppose we are interested in getting x success out of n trials. Now the probability of x successes and consequently (n–x) failures in n independent trials, in a specified order, say, first x trials succes and the remaining (n–x) trials facture is:

$$P(ss \ldots s\ f\ f \ldots f) = P(s)\ P(s) \ldots P(s)\ P(f) \ldots P(f)$$

$$\because \text{ the trials are independent} = \underbrace{p.p \ldots p}_{\text{x factor}}.\ \underbrace{q.q \ldots q}_{(n-x)\text{ factor}}$$

$$= p^x.q^{n-x}$$

But this is simply one way of getting the probabilety of x successes out of n trials. As we know, x successes in n trials can occur in ncx ways, and the probability for each of these cases is p^x q^{n-x}, Hence by the addition theorem of probability the probability of getting x success in n trials is given by,

$p^xq^{n-x} + p^xq^{n-x} + ... + ... + ... + p^xq^{n-x}$, ncx times = ncx p^2 q^{n-x}

Which is the probability mass function of binomial distribution. Again, x, the number of successes out of n trials can take values 0, 1, 2, ..., ..., ..., n. Therefore, we may define a binomial distribution as follows.

11.4.2 Definition

A random variable X is said to follow binomial distribution if it assumes non-w values and if its p.m.f. is given by,

$$p(x) = p(X = x) \begin{cases} {}^{n}c_x\ p^x\ q^{n-x}, & x = 0, 1, 2, ..., ..., ..., n \\ 0, & \text{otherwise.} \end{cases}$$

The two constante n an p are the parameters of this dist^n. It is also abbreviated as B(n, p).

Note: The probability of getting 0 success. 1 success, ..., ..., ..., n successes in this case, i.e. $nc_o\ p^oq^n$, $nc_1p\ q^{n-1}$, ..., ..., ..., ncn $p^0\ q^x$ ... are the successive terms in the expansion of $(q + p)^n$, a binomial expansion, that is why this distribution is also termed as 'binomial distribution'.

11.4.3 Moments of Binomial Distribution

We are interested mainly in the first two moments of benomial distribution which are obtained as follows.

μ_1' = mean = E(x)

$$= \sum_x xp(x)$$

$$= \sum_{x=0}^{n} x n_{c_x} p^x q^{n-x}$$

$$= \sum_{x=0}^{n} x \frac{n^{n-1}}{x} c_{x-1} p^x q^{n-x} \qquad \because\cdot\ n_{c_x} = \frac{n}{x},\ n-1_{c_x}$$

$$= np \sum_{x=1}^{n} n-1_{c_x} p^{x-1} q^{n-x}$$

$$= np\left[{}^{n-1}C_0 p^0 q^{n-1} + n-1_{c_1} pq^{n-2} + \ldots + {}^{n-1}C_1 p^{n-1} q^0\right]$$

$$= np(q+p)^{n-1}$$

$$= np \qquad \because\cdot\ p+q=1$$

Hence the mean of binomial distribution is n p.

Again, $\mu_1' = E(\lambda^2) = E\{X(x-1)\} + E(x)$

$$E\{x(x-1)\} = \sum_{x=0}^{n} x(x-1) n_{c_x} p^x q^{n-x}$$

$$= \sum_{x=0}^{n} x(x-1) \frac{n(n-1)^{n-2}}{x(x-1)} c_{x-2} p^x q^{n-x}$$

$$= n(n-1)p^2 \sum_{x=0}^{n} n-2c_{x-2} p^x q^{n-x}$$

$$= n(n-1)p^2(q+p)^{n-2}$$

$$= n(n-1)p^2$$

$\therefore$ $E(x^2) = n(n-1)p^2 + np$

$\therefore$ μ_2 = variance = $\mu_2' - \mu_1'^2 = n(n-1)p^2 + np - n^2p^2$

$$= n^2p^2 - np^2 + np - n^2p^2$$

$$= np(1-p)$$

$$= npq.$$

So the variance of binomial distribution is npq.

Remark: For binomial distribution mean is always greater than variance. This is an important characteristic of binomial distribution.

11.4.4 Moment Generating Function of Binomial Distribution

If X ~ B (n, p), then

$$M_x(t) = E(e^{tx}) = \sum_x e^{tx} p(x)$$

$$= \sum_{x=0}^{n} e^{tx}\, {}^nc_x p^x q^{n-x}$$

$$= \sum_{x=0}^{n} {}^nc_x \left(e^t p\right)^2 q^{n-x}$$

$$= {}^nc_0 \left(e^t p\right)^0 q^n + {}^nc_1 \left(e^t p\right) q + \ldots + {}^nc_n \left(e^t p\right)^n q^r$$

$$= (q + e^t p)^n$$

Which is the moment generating function of binomial dist^n.

11.4.5 Fitting of Binomial Distribution

The recurrence relation for probabilities of binomial distribution is,

$$p(x+1) = \left(\frac{n-x}{x+1}\right)\left(\frac{p}{q}\right) p(x) \text{ . } x = 0, 1, 2, \ldots, \ldots, \ldots, n-1$$

Putting x = 0, 1, 2, ..., ..., ..., (n – 1) we get p(1), p(2), ..., ..., ..., p(n). Here, we need to calculate only p(0), which is:

$$p(0) = {}^nc_0 p^0 q^n = q^n$$

For felling a binomial distribution, we need to find the values of n and p. n is generally known, and p = $\frac{np}{n}$.

Where np is the mean of the given set of data.

After calculating the value of p, we calculate the probabilities p(0), p(1), ..., ..., ..., p(n) as discussed above. Now multiplying these probabilities by N, the total frequency, we get the expected frequencies for x = 0, 1, 2, ..., ..., ..., n. Now we shall illustrate the

fitting of biomial distribution with the help of an example.

Note: Calculation of expected frequency is called fitting.

Example: 3 coins are tossed 64 times and the following distribution of number of heads were obtained.

Number of heads:	0	1	2	3
Frequency:	10	18	20	16

Fit a binomial distribution under the hypothesis that the coins are unbiased.

Solution: Since the coins are unbiased

$p = q - \frac{1}{2}$ [in case coins are biased p is to be calculated from mean]

Here, N = 64, n = 3

Now, $p(0) = q^n = \left(\frac{1}{2}\right)^3 = \frac{1}{2^3} = \frac{1}{8}$

Now using recurrence relation, $p(x+1) = \left(\frac{n-x}{x+1}\right)\left(\frac{p}{q}\right)p(x)$

We calculate the probabilities as,

$p(1) = \frac{3-0}{0+1}\frac{p}{q}p(0) = 3.\frac{1}{2^3} = \frac{3}{8}$

$p(2) = \frac{3-1}{1+1}\frac{p}{q}p(1) = \frac{3}{2}.\frac{3}{8} = \frac{3}{8}$

$p(3) = \frac{3-2}{2+1}\frac{p}{q}p(2) = \frac{1}{3}.\frac{3}{8} = \frac{1}{8}$

Hence the expected frequenties are,

x	Expected frequency
0	$n \times p(0) = 64 \times \frac{1}{8} = 8$
1	$n \times p(1) = 64 \times \frac{3}{8} = 24$
2	$n \times p(2) = 64 \times \frac{3}{8} = 24$
3	$n \times p(3) = 64 \times \frac{1}{8} = 8$

11.4.6 Properties of Binomial Distribution

i) The binomial distribution is a discrate distribution where the random variable X takes the values 0, 1, 2, ..., ..., ..., n.

ii) The binomial distribution has two parameters, n and p.

iii) The mean is equal to np and variance is given by npq.

iv) The mean of binomial distribution is always greater than variance.

v) If X_1 and X_2 follows binomial distribution with parameters (n_1, p) and (n_2, p) respectively, these $x_1 + x_2$ follows binomial distribution with parameter ($n_1 + n_2$, p). THis is known as the additive property of binomial distribution.

vi) Binomial distribution with parameters (n.p) tends to poisson distribution if n is very largs, p is very small and np is finite.

vii) Binomial distributions also tends to Normal distribution if n is very large and neither non q is very small.

11.4.7 Importance of Binomial Distribution

Theoretical distribution like Binomial distribution provide data on the basis of which observed (empirical or experimental) results can be assessed. In fact, where theoretical distributions are available one need not have to bother for observed distributions. Theoretical distributions provide the decision maker with a sound basis for taking rational and dependable decisions. However in order to apply Binomial distribution, the assumptions of the distribution must hold true which are: two mutually exclusive outcomes, a fixed probability of outcome in any trial and independent trials.

11.4.8 Illustrative Examples

1) A r. v. X follows binomial distribution with mean $\frac{5}{3}$ and P(x = 2) = P (x = 1). Find variance, P(x = at leart 1), P(x = at most 1).

Solution: If X follows binomial distribution with parameters n an p, the p.m.f. of X is,

$p(x) = {}^{n}c_x p^x q^{n-x}$, x = 0, 1, 2, ..., ..., ..., n; q = 1 – p

Given, mean = np = $\frac{5}{3}$

and P(x = 2) = P(x = 1)

$\Rightarrow {}^{n}c_2 p^2 q^{n-2} = {}^{n}c_1 pq^{n-1}$

$\Rightarrow \frac{n(n-1)}{2}\frac{p}{q} = n$

$\Rightarrow (n-1)p = 2q$

$\Rightarrow np - p = 2q$

$\Rightarrow \frac{5}{3} - p = 2q$

$\Rightarrow 5 - 3p = 6q$

$\Rightarrow 6q + 3p = 5$

$\Rightarrow 6(1-p) + 3p = 5$

$\Rightarrow 6 - 6q + 3p = 5$

$\frac{5}{3}$

$\Rightarrow -3p = -1$

$\Rightarrow p = \frac{1}{3}$

$\therefore q = 1 - p = \frac{2}{3}$

Again, $np = \frac{5}{3}$ $\quad\quad$ $\therefore n = \frac{\frac{5}{3}}{\frac{1}{3}} = 5$

Hence, variance = $npq = 5.\frac{1}{3}.\frac{2}{3} = \frac{10}{9}$

P(x = at least 1) = $p(x \geq 1)$

= 1 – p(x = 0) $\quad \because p(x = 0) + (x \geq 1) = 1$

$= 1 - {}^{5}C_0\left(\frac{1}{3}\right)\left(\frac{2}{3}\right)^5$

$= 1 - \left(\frac{2}{3}\right)^5$

$= \frac{211}{243}$

$$P(x = \text{at most } 1) = P(x = 0) + p(x = 0)$$

$$= {}^{5}C_{0}\left(\frac{1}{3}\right)^{0}\left(\frac{2}{3}\right)^{5} + {}^{5}C_{1}\left(\frac{1}{3}\right)^{1}\left(\frac{2}{3}\right)^{4}$$

$$= \frac{112}{243}$$

2) Probability of follows in statistics examination is 15%, 20 batchs of 5 students each appear for the examination. Obtain in how many batches 3 or more students would pass.

Solution: Given q = probability of failure = 15% = 0.15

$\therefore$ q = probability of success = 0.85

n = 5.

If X denote the number of students who would pass, then

$p(x) = {}^{n}c_{x}p^{x}q^{n-x}$, x = 0, 1, ..., ..., ..., 5.

We need $p(x \geq 3) = p(x = 3) + p(x = 4) + p(x = 5)$

$= {}^{5}C_{3}p^{3}q^{5-3} + {}^{5}C_{4}p^{4}q^{5-4} + {}^{5}C_{5}p^{5}q^{0}$

$= 10 \times (.85)^{3}(.15)^{2} + 5(0.5)^{4}(0.15) + (0.15)^{5}$

$= 0.1381 + 0.3915 + 0.4437$

$= 0.9733.$

$\therefore$ Required number of batches in which 30x more students would pass = 20 x 0.9733

$= 19.468 \approx 19.$



CHECK YOUR PROGRESS

Q.2: If $X \sim B\left(12, \frac{1}{3}\right)$ what is E(x)?

Q.3: Examine the following statement and add your comment, "The mean of a binomial distribution is 5 and s.d. is 3".

Q.4: What are the parameters of binomial distribution?

Q.5: Find the varience of the binomial distribution

$$p(x) = {}^{10}C_{x}\left(\frac{2}{5}\right)^{x}\left(\frac{3}{5}\right)^{10-x}, \quad x = 0, 1, 2, ..., ..., ..., 10.$$

Q.6: Define Bernoulli treats. Find its distributions and also its mean and variance.

Q.7: Derive Binomial distribution.

or

Obtain the probability of exactly x successes in n independent Berioulli treats.

11.5 POISSON PROCESS

The Poisson process was first development by French mathematician and physicist serieon Devis Poisson, who published if in the year 1837. In the class of univaxiate distributions, the poisson distribution is perhaps the most useful one, because of its wids applicability in the description of natural phenomena, for example, demand for service in a big departmental store, major accident rate per day in a certain highway etc. A typical application of poisson distribution is for analysing queing (or waiting live) problems in which arriving customers during an interval of time arrive independently and the number of arrivals depends on the length of the time interval.

11.5.1 Poisson Distribution as a Limiting Case of Binomial Distribution

The binomial probability law of rare events gives rise to poisson probability law. That is poisson distribution can be derived as a limiting case of binomial distribution under the following conditions.

i) n, the number of independent Bernoulli trials is very large, i.e. $n \to \alpha$

ii) p, the probability of success in each trial is very small, i.e. $p \to \alpha$

iii) $np = \lambda$ (say), is finits, where λ is a positive real number. under the above conditions, we can derive the p.m.f. of poisson distribution as follows:

Let us consider the p.m.f. of poisson distribution as,

$$p(x) = \binom{n}{x} p^x q^{n-x}, \quad x = 0, 1, 2, ..., ..., ..., n; \; q = 1 - p.$$

$$\text{Now, } \lim_{n\to\alpha} p(x) = \lim_{n\to\alpha} \binom{n}{x} p^x q^{n-x}$$

$$= \lim_{n\to\alpha} \frac{n!}{x!(n-x)!}\left(\frac{\lambda}{n}\right)^x \left(1-\frac{\lambda}{n}\right)^{n-x} \quad \because np = \lambda$$

$$= \frac{\lambda^x}{n!} \lim_{n\to\alpha} \frac{n!}{(n-x)!n^x}\left(1-\frac{\lambda}{n}\right)^{n-x}$$

$$= \frac{\lambda^x}{n!} \lim_{n\to\alpha} \frac{n!}{(n-x)!n^x} \lim_{n\to\alpha}\left(1-\frac{\lambda}{n}\right)^{n} \lim_{n\to\alpha}\left(1-\frac{\lambda}{n}\right)^{-x}$$

$$= \frac{\lambda^x}{n!} \lim_{n\to\alpha} \frac{n(n-1)(n-2)...(n-\overline{x-1})(n-x)!}{(n-x)!n^x} \quad \therefore e^{-\lambda}$$

$$= \frac{\lambda^x}{n!} \lim_{n\to\alpha} \frac{n^x\left\{1.\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)...\left(1-\frac{x-1}{n}\right)\right\}}{n^x}, e^{-\lambda}$$

$$= \frac{e^{-\lambda}\lambda^x}{x!} \quad \because \lim_{n\to\alpha} \frac{1}{n} \to 0$$

Which is the p.m.f of poisson distribution. We give below the formal definition of poisson distribution as,

Definition: A discrate random variable X is said to follow poisson distribution, if it assumes non-negative values only and its p.m.f is given by,

$$p(x) = \frac{e^{-\lambda}\lambda^x}{x!}, \; x = 0, 1, 2, ..., ..., ..., \lambda$$

= 0, else where.

λ is the only parameter of this distribution.

11.5.2 Moments of Poisson Distribution

The arithmatic mean of poisson distribution is given by,

$$\mu_1' = E(x) = \sum_x xp(x)$$

$$= \sum_{x=0}^{\alpha} x\frac{e^{-\lambda}\lambda^x}{x!}$$

$$= e^{-\lambda}.\lambda \sum_{x=1}^{\alpha} \frac{x\lambda^{x-1}}{x(x-1)!}$$

$$= e^{-\lambda}.\lambda \sum_{x=1}^{\alpha} \frac{\lambda^{x-1}}{(x-1)!}$$

$$= e^{-\lambda}.\lambda \left[1 + \lambda + \frac{\lambda^2}{2!} + \ldots\right]$$

$$= e^{-\lambda}.\lambda.e^{\lambda}$$

$$= \lambda$$

$\therefore$ the mean of poisson distribution is λ.

$\mu_2' = E(x^2) = E\{x(x-1)\} + E(x)$

Now, $= E\{x(x-1)\}$

$$= \sum_{x} x(x-1)p(x)$$

$$= \sum_{x=0}^{\alpha} x(x-1)\frac{e^{-\lambda}\lambda^x}{x!}$$

The exponential exponsion,

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots$$

$$= e^{-\lambda}.\lambda^2 \sum_{x=2}^{\alpha} x(x-1)\frac{\lambda^{x-2}}{x.(x-1)(x-2)!}$$

$$= e^{-\lambda}.\lambda^2 \sum_{x=2}^{\alpha} \frac{\lambda^{x-2}}{(x-2)!}$$

$$= e^{-\lambda}.\lambda^2 \left[1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \ldots\right]$$

$$= e^{-\lambda}.\lambda^2.e^{\lambda}$$

$$= \lambda^2$$

$\therefore E(\lambda^2) = \lambda^2 + \lambda$

Now, $\nu(x) = \mu_2 = \mu_2' - \mu_1'^2$

$= \lambda^2 + \lambda - \lambda = \lambda$

Hence the variance of poisson distribution is also λ.

Remark: Poisson distribution is the only discrate distribution where mean is equal to variance.

11.5.3 Moment Generating Function

The moment generating function of poisson distribution is given by,

$$mx(t) = E(e^{tx}) = \sum_{x=0}^{\alpha} e^{tx} . \frac{e^{-\lambda}\lambda}{x!}$$

$$= e^{-\lambda} \sum_{x=0}^{\alpha} \frac{\left(e^{t}\lambda\right)^2}{x!}$$

$$= e^{-\lambda} \left[1 + e^{t}\lambda + \frac{\left(e^{t}\lambda\right)^2}{2!} + \frac{\left(e^{t}\lambda\right)^3}{3!} + \ldots \right]$$

$$= e^{-\lambda} e^{e^{t}\lambda}$$

$$= e^{-\lambda}(1 - e^{t})$$

11.5.4 Fitting of Poisson Distribution

The recurrence relation for probabilities of poisson distribution is,

$$p(x+1) = \frac{\lambda}{x+1} p(x), \; x = 0, 1, 2, \ldots, \ldots, \ldots \qquad (1)$$

Putting x = 0, 1, 2, ..., ..., ... in (1) we get the probabilities p(1), p(2), ..., ..., ... etc. so we need to calculate p(0) only, which is,

$$p(0) = \frac{e^{-\lambda}\lambda^{0}}{0^{0}} = e^{-\lambda}$$

where, λ is the mean of poisson distribution is to be obtained from data to be fitted.

Ultimately, the expected frequencies are calculated by multiplying the corresponding probabilities by N, the total frequency. We shall explain the fitting of poisson distribution to a given data set.

Example: 200 drivers made the following road accidents

Accidents:	0	1	2	3	4
No. of drivers:	122	60	15	2	1

Fit a theoretical distribution to the data.

Solution: The suitable theoretical distribution to be fitted here may be the poission distribution,

Let, x_i = number of road accidents

f_i = no. of drivers.

Total number of drivers = N = 200

Here, λ = mean of poisson distribution

$$= \frac{1}{N}\sum_{i=1}^{n} fixi$$

$$= \frac{100}{200}$$

$$= 0.5$$

$\therefore$ $p = E^{-0.5} = 0.60653$

Using recurrence relation we get,

$$p(1) = \frac{\lambda}{0+1}p(0) = \lambda.p(0) = 0.5 \times 0.60653 = 0.30327$$

$$p(2) = \left\{e^{-1} + e^{-1} + \frac{e^{-1}}{2!}\right\} p(1) = \frac{0.5}{2} \times 0.60653 = 0.07582$$

$$p(3) = \frac{\lambda}{3}p(2) = \frac{0.5}{3} \times 0.60653 = 0.01264$$

$$p(4) = \frac{\lambda}{4}p(3) = \frac{0.5}{4} \times 0.60653 = 0.00158$$

$\therefore$ the expected frequencies are,

$f(0) = N \times p(0) = 200 \times 0.60653 = 121.306 \approx 121$

$f(1) = N \times p(1) = 200 \times 0.30327 = 60.654 \approx 61$

$f(2) = N \times p(2) = 200 \times 0.07582 = 15.164 \approx 15$

$f(3) = N \times p(3) = 200 \times 0.01264 = 2.528 \approx 3$

$f(4) = N \times p(4) = 200 \times 0.00158 = 0.316 \approx 0$

Total frequency = 200

11.5.5 Properties of Poisson Distribution

i) Poisson distribution is a discreta distribution where the random variable X lakes the values 0, 1, 2, ..., ..., ..., λ.

ii) The poisson distribution has only one parameter λ.

iii) The mean and the variance of poisson distribution are equal and is equal to λ.

iv) The poisson distribution has only one or two modes.

v) Poisson distribution can be obtained as a lemiting case of beniomial distribution.

vi) If X and Y are independent poisson variates with paremeters λ_1 and λ_2 respectively, then x + y is also a poisson variate with parameter $\lambda_1 + \lambda_2$.

11.5.6 Application of Poisson Distribution

Physical situations in daily life where poisson distribution is applicable is

i) Number of telephone calls recieved per minute at a particular switch board.

ii) Number of accidents in a city per unit time.

iii) Number of defective parts produced in a factory per unit time.

iv) Number of printing mistakes in each page of a book.

11.5.7 Illustrative Examples

Example 1: Find the mean and variance of poisson distributionsuch that, P(x = 2) = P(x = 1)

Solution: It X follows poisson distribution with parameter λ,

then, $p(x) = \frac{e^{-\lambda}\lambda^x}{x!}$, x = 0, 1, 2, ..., ..., ...

Now, p(x = 2) = p(x = 1)

$$\Rightarrow \frac{e^{-\lambda}\lambda^2}{2!} = \frac{e^{-\lambda}.\lambda}{1!}$$

$$\Rightarrow \frac{\lambda}{2} = 1$$

$$\Rightarrow \lambda = 2$$

$\therefore$ the man and variance of poisson distribution is equal to $\lambda = 2$.

Example 2: The number of teckellers travallers deteched per checking follows poisson distribution with mean 10. Find the probability of getting two ticketters tracullers.

Solution: Let X denote the no. of ticketters traceallers in a checking. so X follows poisson distribution with given mean.

$\lambda = 10$

$\therefore p(x) = \frac{e^{-10}10^x}{x!}$, $x = 0, 1, 2, ..., ..., ...$

$\therefore$ P(getting two ticketters tracullers during a checking)

$= P(x = 2)$

$= \frac{e^{-10}10^2}{2!}$

$= 0.0025.$

Example 3: What probability model is appropriats to describes a situation where 100 misprints are distributed randomly throughout the 100 pages of a book? For this model what is the probability that a page observed at random will contain at least 3 misprients.

Solution: Since 100 misprints are distrebuted randomly throughout the 100 pages of the book, the probability that there being a mesprint is $p = \frac{1}{100} = 0.01$, which is very small.

n = 100 pages is quite large.

Hence poisson distribution is but reaited in this case.

Here, $\lambda = np = 0.01 \times 100 = 1$

$\therefore$ Probability that there are alteast three misprints in a page is:

$p(x \geq 3) = 1 - p(x < 3)$

$= 1 - \{P(x = 0) + P(x = 1) + P(x = 2)$

$= 1 - \left\{e^{-1} + e^{-1} + \frac{e^{-1}}{2!}\right\}$

$= 1 - 2.5e^{-1}$

$= 1 - 2.5 \times 0.3679$

$= 0.0802$



CHECK YOUR PROGRESS

Q.8: Fill in the blanks:

i) For a poisson distribution mean is ..... to variance (less/equal/greater)

ii) The s.d. of a poisson distribution with mean 4 is ..... (4/2).

iii) For a poisson distribution $\frac{e^{-1.5}(1.5)^x}{x!}$, x = 0, 1, 2, ... the varience is (1.5/2.5).

iv) If x follows poisson distribution is ..... $\frac{e^{-\lambda(1-e^t)}}{e^{-\lambda}(1-2)^t}$.

v) If x_i, i = 1, 2, ..., ..., ..., n are n poisson variates then the distribution of $\sum_{i=1}^{n} x_i$ is also a ..... variate. (Binomail/Poisson)

Q.9: "Poisson distribution is also known as probability distribution of rare events." –Explain.

Q.10: Give some examples of poisson distribution.



11.6 LET US SUM UP

- A function, which is used to generate moments of a probability distribution is known as moment generating function.
- The importance of moment generating function in the fact that, a probability distribution can be completely determined if its moments are known.
- The moment generating function of a random variable X is given by, $Mx(t) = E(e^{tx})$, $t > 0$.
- A random experiment whose outcoms are classified into two categories, called 'success' and 'failure' is called a Bernoulli treats.

- The p.m.f. of binomial distribution with parameters n and p is, $p(x) = {}^nC_x p^x q^{n-x}$, x = 0, 1, 2, ..., ..., ..., n.
- The mean of binomial distribution is np and varience is npq.
- For a binomial distribution, mean is always greater than variance.
- Sum of independent binomial variates is also a binomial variate.
- Poisson distribution can be derived as a limiting case of benomial distribution when, $n \to \alpha$, $p \to o$ and $np \to \lambda$ (say).
- The p. m. f. of poisson distribution with parameter λ is given by,

  $p(x) = \frac{e^{-\lambda}\lambda^x}{x!}$, x = 0, 1, 2, ..., ..., ... λ.
- Poisson distribution plays an important role in the discription of random phenomena in a continuous interval of any kind. For example, telephone calls at a telephone exchange within a specified period of time.
- Sum of independent poisson variates is also a poisson variate.



11.7 FURTHER READING

1) Das, K. K, Bhattacharjee, D. (2008). *A Treatere on Statistical Inference and Distributions*. Asian Books Pvt. Ltd.
2) Gupta, S. C. and Kapoor, V. K. (2002). *Fundementals of Mathematical Statistics*. Sultan Chand & Sons.
3) Goon, A. M., Gupta. M. P. and Dasgupta, B. (2002). *Fundementals of Statistics* (Vol. I). Seventh Revised Edition. Kolkata: The World Press Pvt. Ltd.
4) Saxena, H. C. (1994). *Probability and its Application*. Sixth Edition. New Delhi: S. Chand and Company.
5) Singh, R. (2012). *An Introduction to Probability and Probability Distributions*. Books and Ellied (P) Ltd.

11.8 ANSWERS TO CHECK YOUR PROGRESS

Ans. to Q. No. 1: i) True, ii) False, iii) False, iv) True.

Ans. to Q. No. 2: We know that if X ~ B (n, p) then E(x) = np.

Here, n = 12, = $\frac{1}{3}$

$\therefore$ E(x) = np = 12. $\frac{1}{3}$ = 4

Ans. to Q. No. 3: Here, given, mean = 5, s.d. = 3

$\therefore$ variance = 9

But we know that for a benomial distribution, mean is always greater than variance. Hences, the above statement is false.

Ans. to Q. No. 4: The parameters of a binomial distribution are n and p.

Ans. to Q. No. 5: Given, $p(x) = {}^{10}C_x\left(\frac{2}{5}\right)^2\left(\frac{3}{5}\right)^{10-x}$, x = 0, 1, 2, ..., ..., ..., 10.

Here, n = 10, p = $\frac{2}{5}$, q = $\frac{3}{5}$

Hence, variance = npq = 10. $\frac{2}{5}.\frac{3}{5}$ = $\frac{12}{5}$

Ans. to Q. No. 6: A particular trial having two outcomes only, viz success and failure occurring with probability p and (1–p) respectively is called a Bernoulli trial. e.g. tossing of a coin.

A random variable X is said to follow Beruoulli distribution if its p.m.f. is given by, $p(x) = p^x(1-p)^{1-x}$, x = 0, 1.

P is the only parameter of this distribution

$\therefore$ Mean = E(x) = 0.(1–p) + 1.p = p.

$V(x) = E(x^2) - \{E(x)^2\}$.

$E(x^2) = 0^2.(x-p) + 1^2.p = p$

$\therefore$ $V(x) = p - p^2 = p(1-p)$.

Ans. to Q. No. 7: Consider n independent Bernoulli trials each having two and two outcomes only called success and failure, Probability of success is p and the probability of failure is (1 – p) = q.

The probability of x successes and (n–x) failures in n independent trials in a specific order, say, s s s... f.f.f is given by,

$$P(s.\ s.\ .\ .\ .\ s.\ f\ .\ f.\ .\ .\ .\ f) = p(s)\ p(s)\ s\ .\ .\ .\ p(s)\ p(f)\ .\ .\ .\ p(f)$$

$$\because \text{trials are independent} = \underbrace{p.p\ .\ .\ .\ p}_{x \text{ times}}\ \underbrace{q.q\ .\ .\ .\ q}_{(n-x) \text{ times}}$$

$$= p^x q^{n-x}$$

But, x successes in n trails can occur in n_{CX} .... which are mutually exclusive and the probability for each of these is $p^x\ q^{n-x}$. Hence by addition theorem of probability,

$$p(x = x) = n_{CX}\ p^x\ q^{n-x},\ x = 0, 1, 2, .\ .\ ., n.$$

Ans. to Q. No. 8: i) equal, ii) 2, iii) 1.5, iv) $e^{-\lambda(1-e^t)}$, v) poisson.

Ans. to Q. No. 9: Poisson distribution is also known as probability distribution of rare events, because poisson distribution is a positivaly skewed distribution, i.e probabilities tend to zero for large number of occurrences. For these poisson distribution is also known as probability distribution of rare events.

Ans. to Q. No. 10: Some examples of poisson distribution are,

i) number of deaths from suicide reported in a particular city.

ii) number of defictive items in a lot of manifactures items.

iii) number of printing mistakes in each page of a book.



11.9 MODEL QUESTIONS

Q.1: Define moment guerating function and state its properties.

Q.2: Write a note on binomial distribution.

Q.3: Define a binomial distribution distribution and also obtain its mean and variance.

Q.4: Obtain poisson distribution as a limiting case of binomial distribution.

Q.5: If X follows poisson distribution with $V(\lambda) = 2$, find its p.m.f.

Q.6: Define poisson distribution, state the conditions under which the binomial p.m.f. can be approximated by the poisson p.m.f.

Q.7: Obtain m.g.f. of poisson distribution and hence find its mean.

*** ***** ***

UNIT 12: THEORETICAL DISTRIBUTIONS (CONTINUOUS VARIABLE)

UNIT STRUCTURE

12.1 Learning Objectives

12.2 Introduction

12.3 Continuous Probability Distribution

12.4 Normal Distribution

12.4.1 Properties of Normal Distribution

12.4.2 Standard Normal Variate

12.5 Let Us Sum Up

12.6 Further Reading

12.7 Answers To Check Your Progress

12.8 Model Questions

12.1 LEARNING OBJECTIVES

After going through this unit, you will able to:

- continuous probability distribution and probability deusity function
- normal distribution
- standard normal variate
- properties of normal distribution.

12.2 INTRODUCTION

Normal distribution is a continuous distribution. It is a limiting case of the binomial distribution. It has immense application. It is the most popular and commonly used distribution. It was discovered by Abrahoun de Moivre (1667–1754) in 1733.

12.3 CONTINUOUS PROBABILITY DISTRIBUTION

If a random variable is a continuous variable, its probability distribution is called a continuous probability distribution.

A continuous probability distribution cannot be expresed in tabular form. An equation or formula is used to describe a continuous probability

distribution. The equation used to describe a continuous probability distribution is called a probability. density function (p.d.f.)

12.4 NORMAL DISTRIBUTION

Normal distribution is derived from the binomial distribution under two conditions, viz.

i) n, the number of trials, is sufficiently large, i.e. $n \to \infty$

and ii) neither p nor q is very small.

A random variable X is said to follow normal distribution, if and only if, its probability density function is given by

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

Where x is the real value of X, i.e. $-\infty < X < \infty$

The variable X is said to be distributed normally with mean μ and variance α^2 symbolically we write $X \sim N(\mu, \alpha^2)$. It is read as 'X is distributed as $N(\mu, \alpha^2)$'.

Note:

1) The probability density function given by (1) has two parameters, viz., μ and α.

   μ is any real number, i.e. $-\infty < \mu < \infty$ whereas α is a positive real number, i.e., $\alpha > 0$.

2) We know that probability can never be negative.

   $\therefore f(x) \geq 0, \ \forall x$

3) If $\mu = 0$, $\alpha = 1$ the p.d.f is given by

$$f(x) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}$$

In this case $X \sim N(0, 1)$. In this case X is called the standard normal variate, and the p.d.f given by (2) is called standard normal distribution.

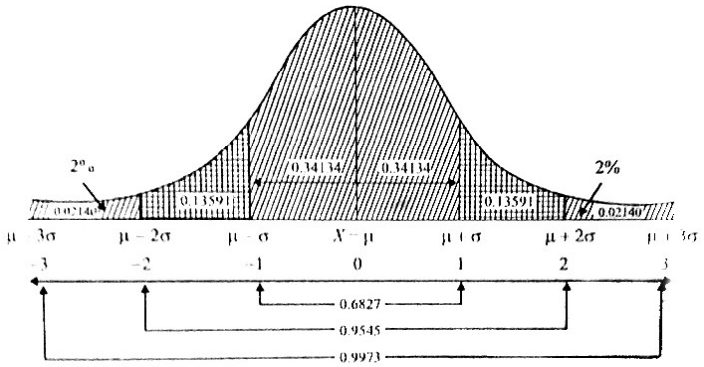
12.4.1 Properties of Normal Distribution

Normal distribution has many properties, some of which are given below.

1) The equation of the normal probability curve is:

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

It is of the shape given below. It is bell-shaped.



[*Source:* **Biostatistics** by Dr. P. N. Arora and Dr. P. K. Malhan (Himalya Publishing House, 2010)]

2) The curve is symmetrical about the line x = ∞, and x lies between $-\infty$ and $+\infty$.

3) At x = ∞, the ordinate is maximum, and it is $Y = \frac{1}{\sigma\sqrt{2\pi}}$.

4) The total area under the normal curve within the limits $-\infty$ to ∞ is 1, i.e.,

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

5) The percentage distribution of area under the normal curve given below, and is shown in the figure given above.
   i) About 68% of the area falls between $\mu - \alpha$ and $\mu + \alpha$.
   ii) About 95.5% of the area falls between $\mu - 2\alpha$ and $\mu + 2\alpha$.
   iii) About 99.7% of the area falls between $\mu - 3\alpha$ and $\mu + 3\alpha$.

6) The curve has a single peak, i.e., it is unimodal.

7) The distribution is symmctrical. Mean = median = mode = μ.

8) The curve is asymptotic, i.e., it meets the x-axis at both the ends at infinity. Theeretically the curve never towches the x-axis.

9) The points of inflexion are x = $\mu - \alpha$ and x = $\mu + \alpha$. [Inflexion points are where the function changes concavity.]

10) Coefficient of skewness = 0, and kurtosis = 3.

11) Mean deviation about mean $\simeq \frac{4}{5}\alpha$.

12) Q.D. $\simeq \frac{2}{3}\alpha$

12.4.2 Standard Normal Variate

Let a random variable X be distributed as $N(\mu, \alpha^2)$. The location and shape of the normal curve depends on μ, and α, Where $-\infty < \mu < \infty$, and $\alpha > 0$. Therefore, no master table for the area can be prepared. This hindrance can be removed if we introduce a new variable Z defined by.

$$Z = \frac{X - \mu}{\alpha}$$

In this case Z is called the standard normal variate.

Now we prove that the mean of Z is 0, and the standard deviation is 1. Before that we state some related properties.

We know that mean of a random variable X is given by

mean = $E(X) = \Sigma x_i p_i$

We prove that–

i) $E(a) = a$; a is a constant

ii) $E(aX + b) = aE(x) + b$; a, b are constants

iii) $E(X - \bar{X}) = 0$

Proof: i) $E(a) = \Sigma a p_i$

$= a\Sigma p_i$

$= a,$ $\because \Sigma p_i = 1$

ii) $(aX + b) = \Sigma(ax_i + b) = \Sigma(ax_i + b)p_i$

$= a\Sigma x_i p_i + b\Sigma p_i$

$= aE(X) + b,$ $\because \Sigma p_i = 1$

iii) $E(X - \bar{X}) = E(X) - E(\bar{X})$

$= \bar{X} - \bar{X},$ $\because \bar{X}$ is constant

$= 0$

Also, we note that–

iv) Var (aX) = a^2 Var (X), where a is a constant

v) Var (aX $\pm$ b) = a^2 Var (X), where a, b are constants.

Now we prove that the mean of Z is 0, and the standard deviation is 1.

Mean = E (Z)

$$= E\left(\frac{X-\mu}{\alpha}\right)$$

$$= \frac{1}{\alpha}E(X-\mu), \qquad \text{property (ii)}$$

$$= \frac{1}{\alpha}[E(X)-\mu], \qquad \text{property (iii)}$$

$$= \frac{1}{\alpha}[\mu-\mu]$$

$$= 0$$

Variance = Var (Z)

$$= \text{Var}\left(\frac{X-\mu}{\alpha}\right)$$

$$= \frac{1}{\alpha^2}\text{var}(X-\mu), \qquad \text{property (v)}$$

$$= \frac{1}{\alpha^2}\text{var } X, \qquad \text{property (iv)}$$

$$= \frac{1}{\alpha^2}.\alpha^2$$

$$= 1$$

Thus, $Z = \frac{X-\mu}{\alpha}$ is distributed as N (0, 1), i.e., Z ~ N(0, 1).

All the properties for a normal variable hold good for standard normal variable (μ = 0, α = 1).

Let Z be a standard normal variate N(0, 1). The area under the standard normal curve is given by a table which is addended at the end of this chapter.

We know that the total area under the normal curve within the limits $-\infty$ to ∞ is 1. The standard normal curve is symmetrical

about Z = 0. Therefore, the area to the left of z = 0 is 0.5, and the area to the right of z = 0 is 0.5.

Suppose, we are to find the area between z = 0, and z = 1.89, i.e, we are to find out $P(0 \leq Z \leq 1.89)$.

First we find 1.8 at the left of the table and single out that row.

Next we find 0.09 at the top of the table and single out that column.

Then we obtain the entry where the row for 1.8 and the column for 0.09 meet. the entry is 0.47062.

Thus $P(0 \leq Z \leq 1.89)$

= (the are from z = 0 to z = 1.89)

= 0.47062.

Example 1: Find the probability of the standard normal variate z ehen z is greater than 1.09.

Solution: We know that the total area under the normal curve within the limits $-\infty$ to ∞ is 1. Therefore, the area to the right of z = 0 is 0.5, as the curve is symmetrical about z = 0.

$\therefore P(Z > 1.09)$

= 0.5 – 0.36214 = 0.13786

Example 2: Find the probability of the standard normal variate z when z is less than 1.09.

Solution: We know that the total area under the normal curve within the limits $-\infty$ to $+\infty$ is 1. Therefore, the area to the left of z = 0 is 0.5, and the area to the right or z = 0 is 0.5.

Again, from the table, we get that the area from z = 0 to z = 1.09 is 0.36214.

$\therefore P(z < 1.09)$

= (area to the left of z = 0) + (area from z = 0 to z = 1.09)

= 0.5 + 0.36214

= 0.86214

Example 3: Find $P(0.5 < z < 1.09)$

Solution: We know that the area to the right of z = 0 is 0.5. From the table, the area from z = 0 to z = 0.5 is 0.19146. Again, the area from z = 0 to z = 1.09 is 0.36214.

$\therefore$ P(0.5 < z < 1.09)

= (Area from z = 0 to z = 1.09) – (area from z = 0 to x = 0.5)

= 0.36214 – 0.19146

= 0.17068

Example 4: A large number of measurements is normally distributed with mean 65.5 cm and standard deviation 6.2 cm. Find the percentage of measurements that fall between 54.8 cm and 68.8 cm.

Solution: Let X be the normal variate showing the large number of measurements.

We have μ = 65.5, α = 6.2

Now, $Z = \frac{X - \mu}{\alpha} = \frac{X - 65.5}{6.2}$

When, X = 54.8 $Z = \frac{54.8 - 65.5}{6.2} = \frac{-10.7}{6.2} = -1.7$

When, X = 68.8, $Z = \frac{68.8 - 65.5}{6.2} = 0.5$

Thus, when X lies between 54.8 and 68.8, Z lies between –1.7 and 0.5

Now P(–1.7 < X < 0.5)

= P(–1.7 < Z < 0) + P(0 < Z < 0.5)

= 0.45543 + 0.19146

= 0.64689 = 64.69%

$\therefore$ the percentage of measurements is 64.69%

Example 5: For the distribution of makrks of candidates in an examination, mean is 14 and standard deviation is 2.5. Assuming the normality of the distribution, find the probability that a randomly selected student will score above 15.

Solution: Let X be the normal variate showing the scores of candidates.

We have μ = 14, α = 2.5

Now, $Z = \frac{X - \mu}{\alpha} = \frac{X - 14}{2.5}$

When X, $Z = \frac{15 - 14}{2.5} = \frac{1}{2.5} = 0.4$

$\therefore$ P(x > 15) = P(z > 0.4)

= 0.5 – (Area between z = 0 and Z = 0.4)

= 0.5 – 0.15542

= 0.34458.



CHECK YOUR PROGRESS

Q.1: Find: (i) P (Z > 2.58), (ii) P(Z < 1.96), (iii) P (Z > –3)

Q.2: For a certain normal distribution, mean is 50, and standard deviation is 5. Find P($45 \leq X \leq 55$), X being a normal variate.

Q.3: A normal variate X has mean 100 and standard deviation 4. Find the probability that X lies between 92 and 10.8

Q.4: The distribution of scores of students in an examination is normal. If the mean is 45, and the standard deviation is 5, find the percentage of students who scored between 40 and 50.

Q.5: The marks obtained in a certain examination follow normal distribution with mean 45, and standard deviation 10. If 1000 students appear at the examination, calculate the number of students scoring (i) less than 30 marks, and (ii) more thaw 80 marks.

12.5 LET US SUM UP

- Normal distribution is a continous distribution. It is a limiting case of the binomial distribution.
- Normal distribution is derived from the binomial distribution under two conditions, viz.,
  - i) n, the number of trials, is sufficiently large, i.e., $n \to \propto$ and
  - ii) neither p nor q is very small.

  A random variable X is said to follow normal distribution, if and only if, its probability density function is given by

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

Where x is the real value of X, i.e., $-\infty < X < \infty$

The variable X is said to be distributed normally with mean μ and variance α^2. Symbolically, we write $X \sim N(\mu, \alpha^2)$. It is read as 'X' is distributed as $N(\mu, \alpha^2)$.

- The normal curve is bell-shaped, and the curve is symmetrical about the line $x = \mu$, and x lies between $-\infty$, and $+\infty$.
- The total area under the normal curve within the limits $-\infty$ to ∞ is 1.
- If a new variable Z is defined by $Z = \frac{X-\mu}{\alpha}$, Z is called the standard normal variate. The mean of Z is 0, and the standard deviation is 1.

12.6 FURTHER READING

1) Das, K. K, Bhattacharjee, D. (2008). *A Treatere on Statistical Inference and Distributions*. Asian Books Pvt. Ltd.
2) Gupta, S. C. and Kapoor, V. K. (2002). *Fundementals of Mathematical Statistics*. Sultan Chand & Sons.
3) Goon, A. M., Gupta. M. P. and Dasgupta, B. (2002). *Fundementals of Statistics* (Vol. I). Seventh Revised Edition. Kolkata: The World Press Pvt. Ltd.
4) Kakoti, S. C. (2003). *Mathematical Statistics: Theory and Applications. Dibrugarh*, Assam: Kasturba Prakashans.
5) Saxena, H. C. (1994). *Probability and its Application*. Sixth Edition. New Delhi: S. Chand and Company.

12.7 ANSWERS TO CHECK YOUR PROGRESS

Ans. to Q. No. 1: i) We know that tha total area under the normal curve within the limits $-\infty$ to $+\infty$ is 1. Therefore, the area to the left of z = 0 is 0.5, and the area to the night of z = 0 is 0.5. Again, from the table, we get that the area from z = 0 to z = 2.58 is 0.49506

$\therefore$ P (z > 2.58)

= (area to the right of z = 0) – (area from z = 0 to z = 2.58)

= 0.5 – 0.49506

= 0.00494

ii) We know that the total area under the normal curve within the limits $-\infty$ to $+\infty$ is 1. Therefore, the area to the left of z = 0 is 0.5, and the area to the right of z = 0 is 0.5. Again from the table, we get that the area from z = 0 to z = 1.96 is 0.47500.

$\therefore$ P (z < 1.96)

= (area to the left of z = 0) + (area from z = 0 to z = 1.96)

= 0.5 + 0.47500

= 0.97500

iii) We know that the total area under the normal curve within the limits $-\infty$ to ∞ is 1. Therefore, the area to the left of z = 0 is 0.5, and the area to the right of z = 0 is 0.5.

Now, P(Z > –3)

= (area from z = –3 to z = 0) + (area to the right of z = 0)

= (area from z = 0 to z = 3) + (area to the right of z = 0)

= 0.49865 + 0.5

= 0.99865

Ans. to Q. No. 2: X is a normal variate.

We have $\mu = 50$, $\alpha = 5$

Now, $Z = \frac{X - \mu}{\alpha} = \frac{X - 50}{5}$

When X = 45, $Z = \frac{45 - 50}{5} = -1$

When X = 55, $Z = \frac{55 - 50}{5} = 1$

Thus, when X lies between 45 and 55, Z lies between –1 and 1.

Now $P(-1 \le Z \le 1)$

$= P(-1 \le Z \le 0) + P(0 \le Z \le 1)$

= 0.34134 + 0.34134

= 0.68268

Ans. to Q. No. 3: Do yourself.

Ans. to Q. No. 4: Do yourself.

Ans. to Q. No. 5: Let X be the normal variate showing marks.

We have $\mu = 45$, $\alpha = 10$

Now $Z = \frac{X - \mu}{\alpha} = \frac{X - 45}{10}$

When X = 30, $Z = \frac{30 - 45}{10} = \frac{-15}{10} = -1.5$

When X = 80, $Z = \frac{80 - 45}{10} = \frac{35}{10} = 3.5$

i) Now $P(X \leq 30)$

= (area to the left of z = 0) – (area from z = – 1.5 to z = 0)

(area to the left of z = 0) – (area from z = 0) to z = 1.5)

= 0.5 – 0.43319 = 0.06681

$\therefore$ the required number of students

= 1000 x 0.06681

= 67 (approximately)

ii) $P(Z \geq 80)$

$= P(Z \geq 3.5)$

= (area to the right of z = 0) – (area from z = 0 to z = 3.5)

= 0.5 – 0.49977

= 0.00023

Now 1000 × 0.00023 = 0.23

$\therefore$ there are no students scoring more than 80 marks.



12.8 MODEL QUESTIONS

Q.1: Define standard normal variate and write down the probability density function of this variate.

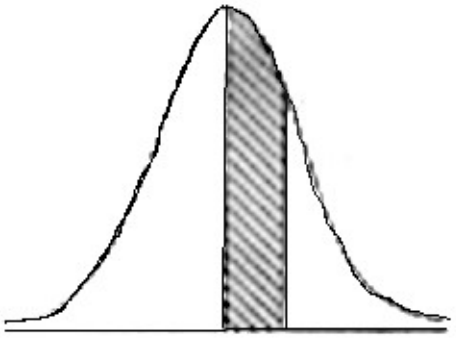
Q.2: State the conditions under which the normal distribution is obtained as a dimiting case of the binomial distribution.

Q.3: Find (i) $P(Z \leq -1.96)$, (ii) $P(Z \leq 1.96)$, (iii) $P(Z > 2.58)$

Q.4: A normal variate X has mean 50 and standard deviation 5. Find $P(40 \leq X \leq 60)$.

Q.5: The mean of a normal distribution is 50 and the standard deviation is 4. What percentage of total frequency lies between 42.16 and 57.84?

Q.6: Let a continuous variable X follow a normal distribution with mean 12 and standard deviation 2. What is the probability that the value of X selected at random lies between 11 and 14?



Area under the Standard Normal Curve

0.0	.00000	.00399	.00798	.01197	.01595	.01994	.02392	.02790	.03188	.03586
0.1	.03983	.04383	.04776	.05172	.05567	.05962	.06356	.06749	.07142	.07535
0.2	.07926	.08317	.08706	.09095	.09483	.09871	.10257	.10642	.11026	.11409
0.3	.11791	.12172	.12552	.12930	.13307	.13683	.14058	.14431	.14803	.15173
0.4	.15542	.15910-	.16276	.16640	.17003	.17364	.17724	.18082	.18439	.18793
0.5	.19146	.19497	.19847	.20194	.20540	.20884	.21226	.21566	.21904	.22240
0.6	.22575	.22907	.23237	.23565	.23891	.24215	.24537	.24857	.25175	.25490
0.7	.25804	.26115	.26424	.26730	.27035	.27337	.27637	.27935	.28230	.28524
0.8	.28814	.29103	.29389	.29673	.29955	.30234	.30511	.30785	.31057	.31372
0.9	.31594	.31859	.32121	.32381	.32639	.32894	.33147	.33398	.33646	.33891
1.0	.34134	.34375	.34614	.34849	.35083	.35314	.35543	.35769	.35993	.36214
1.1	.36433	.36650	.36864	.37076	.37286	.37493	.37698	:37900	.38100	.38298
1.2	.38493	.38686	.38877	.39065	.39251'	.39435	.39617	.39796	.39973	.40147
1.3	.40320	.40490	.40658	.40824	.40988	.41198	.41309	.41466	.41621	.41774
1.4	.41924	.42073-	.42220	.42364	.42507	.42647	.42785	.42922	.43056	.43189
1.5	.43319	.43448	.43574	.43699	.43822	.43943	.44062	.44179	.44295	.44408
1.6	.44520	.44630	.44738	.44845	.44950	.45053	.45154	.45254	.45352	.45449
1.7	.45543	.45637	.45728	.45818	.45907	.45994	.46080	.46164	.46246	.46327

1.8	.46407	.46485	.46562	.46638	.46712	.46784	.46856	.46926	.46995	.47062
1.9	.47128	.47193	.47257	.47320	.47381	.47441	.47500	.47558	.47615	.47670
2.0	.47725	.47784	.47831	.47882	.47932	.47982	.48030	.48077	.48124	.48169
2.1	.48214	.48257	.48300	.48341	.48382	.48422	.48461	.48500	.48537	.48574
2.2	.48610	.48645	.48679	.48713	.48745	.48778	.48809	.48840	.48870	.48899
2.3	.48928	.48956	.48983	.49010	.49036	.49061	.49086	.49111	.49134	.49158
2.4	.49180	.49202	.49224	.49245	.49266	.49286	.49305	.49324	.49343	.49361
2.5	.49379	.49396	.49413	.49430	.49446	.49461'	.49477	.49492	.49506	.49520
2.6	.49534	.49547	.49560	.49573	.49585	.49598	.49609	.49621	.49632.	.49643
2.7	.49653	.49664	.49674	.49683	.49693	.49702	.49711	.49720	.49728	.49736
2.8	.49744	.49752	.49760	.49767	.49774	.49781	.49788	.49795	.49801	.49807
2.9	.49813	.49819	.49825	.49831	.49836	.49841	.49846	.49851	.49856	.49861
3.0	.49865	.49869	.49874	.49878	.49882	.49886	.49889	.49893	.49896	.49900
3.1	.49903	.49906	.49910	.49913	.49916	.49918	.49921	.49924	.49926	.49929
3.2	.49931	.49934	.49936	.49938	.49940	.49942	.49944	.49946	.49948	.49950
3.3	.49952	.49953	.49955	.49957	.49958	.49960	.49961	.49962	.49964	.49965
3.4	.49966	.49968	.49969	.49970	.49971	.49972	.49973	.49974	.49975	.49976
3.5	.49977	.49978	.49978	.49979	.49980	.49981	.49981	.49982	.49983	.49983
3.6	.49984	.49985	.49985	.49986	.49986	.49987	.49987	.49988	.49988	.49989
3.7	.49989	.49990	.49990	.49990	.49991	.49991	.49992	.49992	.49992	.49992
3.8	.49993	.49993	.49993	.49994	.49994	.49994	.49994	.49995	.49995	.49995
3.9	.49995	.49995	.49996	.49996	.49996	.49996	.49996	.49996	.49997	.49997
4.0	.49997	–	–	–	–	–	–	–	–	–
4.5	.499997	–	–	–	–	–	–	–	–	–
5.0	.4999997	–	–	–	–	–	–	–	–	–

[*Source:* Basic statisties by B. L. Agarwal (New Age International Publishers, 2009)]

*** ***** ***

UNIT 13: INDEX NUMBERS

UNIT STRUCTURE

13.1 Learning Objectives

13.2 Introduction

13.3 Uses of Index Numbers

13.4 Types of Simple Index Numbers

13.5 Properties of Relatives

13.6 Construction of Simple Index Numbers

13.7 Construction of Weighted Index Numbers

13.8 Test of Adequacy of Index Numbers

13.9 Let Us Sum Up

13.10 Further Reading

13.11 Answers to Check Your Progress

13.12 Model Questions

13.1 LEARNING OBJECTIVES

After going through this unit, you will be able to:

- discuss about the simple and weighted index numbers
- explain the uses of index numbers
- discuss the different types of index numbers
- explain the properties of different index numbers
- discuss the test of accuracy of index numbers.

13.2 INTRODUCTION

Index number is a statistical device (measure) with a purpose of showing average changes in one or more related variables (like price or quantity) between two periods of time (say, between 1991 and 1996) or two places like cities (say, Delhi and Mumbai) or countries (say, India and Japan). For example, we may be interested in knowing which city of India out of Delhi, Mumbai, Kolkata and Chennai is the costliest or the cheapest in terms of price level. A tourist may be interested in knowing about the cost of

living at different tourist places. Consumer price index number or cost of living index number helps in taking such decisions.

Different commodities are measured in different units. For example, rice and wheat are measured in kilograms, cloth in meters and milk in liters etc. Index numbers attempt some averages relating to commodities which are measured in different units. For purpose of comparison, we are interested in knowing relative changes. These relative changes are expressed in percentage terms.

Index numbers are the indicators of the various trends in an economy. Price index numbers indicate the position of prices whether they are rising or falling and at what rate. Similarly, index numbers regarding agricultural production indicates the trend of change whether it is rising or falling at what rate over a period of time. An index number is an economic data figure reflecting price or quantity compared with a standard or base value. The base usually equals 100 and the index number is usually expressed as 100 times the ratio to the base value. For example, if a commodity costs twice as much in 1970 as it did in 1960, its index number would be 200 relative to 1960. Index numbers are used especially to compare business activity, the cost of living, and employment.

An index number is specialized average. Index numbers may be simple or weighted depending on whether we assign equal importance to every commodities or different importance to different commodities according to the percentage of income spent on them or on the basis of some other criteria. In this chapter, we shall discuss both simple and weighted index numbers.

13.3 USES OF INDEX NUMBERS

In business an index is a statistical measure of changes in a representative group of individual data points.

Index numbers are constructed with various aims and purposes. Index numbers constructed are put into several uses. Some of them are described below:

1) The most commonly used types of index numbers are those that relate to changes in level in prices of a commodity or a group of commodities. Such index numbers are price index numbers. Price index numbers are useful in studying price movements with a view to analyzing their causes, as well as to determine their effects on the economy. To determine economic relationships, it is often useful to compare changes in the market price level with changes in other series, such as gold, bank deposits, bank loans, and the physical volume of production.
2) Indices of physical changes over a period of time in production, in marketing, in sales, in imports and exports, and so on have important uses. Such index numbers over a period of time are known as index time series. These index numbers are extremely useful in the study of the movement of the characteristics over time, in the study of seasonal and cyclic behaviour, and in the study of business cycles, etc.
3) Index numbers can be used for forecasting future trend in demand of commodities or in volume of production etc. Index numbers of industrial and agricultural production not only reflect the trend but also are useful in forecasting production. Index numbers of unemployment which reflect the trend of unemployment are useful in determining factors leading to unemployment and in forecasting future trends.
4) Index numbers which reflect the general price level, and in particular, the consumer price index numbers are useful in determining the quantum of additional wages or dearness allowances to compensate for the changes in the cost of living. The Government of India and most of the State Governments have index linked formulae for grant of dearness allowances to their employees.
5) The changes in consumer price index can be used to determine the real income over a period. Suppose that a person's income in 1980 was 150 times than that in 1970, that is, there is an increase of 50% in money income. Suppose that the cost of living index number in

1980 was 175 times than that in 1970, that is, there is an increase of 75%. Then the person's real income in 1980 is $\frac{150}{175}$ = 85.7% of his income in 1970. Thus index numbers are useful in determining the real income.

13.4 TYPES OF SIMPLE INDEX NUMBERS

1) **Price Relatives:** One of the simplest types of index numbers is a price relative. It is the ratio of the price of a single commodity in a given period or point of time to its price in another period or point of time, called the reference period or base period. If the prices for a period, instead of a point of time, are considered, then suitable price average for the period is taken and these prices are expressed in the same units. If p_0 and p_1 denote the price of a commodity during the base period or reference period (0) and the given period (1) then the price relative of the period 1 with respect to the base period 0 is defined by

Price relative in percentage (of period 1 with respect to 0)

$$= \frac{p_1}{p_0} \times 100 \qquad \ldots \ldots \ldots (7.1)$$

We denote price relative in percentage or without percentage by $\frac{p_1}{p_0}$.

Remark: In the formula of price relative we multiply $\frac{p_1}{p_0}$ by 100 only to get a better expression. This expression is therefore called price relative in percentage.

Exercise 1: If the retail price of fine quality of rice in the year 1980 was Rs.3.75 and that for the year 1983 was Rs.4.50, then find the price relative.

Solution: Here the base period is 1980. The price of rice in the base period was Rs. 3.75. Also in the given period 1983, the price of rice was Rs. 4.50. Using the formula (7.1), the required price relative is:

$$p_{\frac{1980}{1983}} = \frac{\text{Rs. } 4.50}{\text{Rs. } 3.75} \times 100 = 120\%$$

Exercise 2: The exchange rate of a US dollar was Rs. 40.00 in July 2008 and was Rs. 50.00 in February 2009. Find the price relative.

Solution: The price relative of a dollar in February 2009 (say period F) with respect to that in July 2008 (say period J) is given by

$$p_{\frac{F}{J}} = \frac{\text{Rs. } 50.00}{\text{Rs. } 40.00} \times 100 = 125\%$$

2) **Quantity Relatives:** Another simple type of index number is a quantity relative. This is useful when we are interested in changes in quantum or volumes of a commodity such as quantities of production or sale or consumption. Here the commodity is used in a more general sense. It may mean the volume of goods (in tonnes) carried by roadways or the volume of export to a country or import from a country. In such cases, we consider of quantity or volume relatives. If quantities or volumes are taken for a period instead of a point of time, a suitable average is to be taken and the quantities or volumes are to be expressed in the same units. If q_0 and q_1 denote the quantity or volume produced, consumed or transacted during the base period (0) and the given period (1) then quantity relative of the period 1 with respect to the base period 0 is defined by

Quantity relative in percentage (of period 1 with respect to 0)

$$= \frac{q_1}{q_0} \times 100 \qquad \ldots \ldots \ldots (7.2)$$

We denote quantity relative in percentage or without percentage by $q_{\frac{0}{1}}$.

Exercise 3: The production of tea in Assam in the month of January 2009 was 3608 tonnes and that in February 2009 was 3700 tonnes. Find the quantity relative.

Solution: The quantity relative of tea in February 2009 (say period F) with respect to that in January 2009 (say period J) is given by

$$q_{\frac{F}{J}} = \frac{3700}{3608} \times 100 = 102.55\%$$

3) Value Relatives: A value relative is another type of simple index number. It is usable when we wish to compare changes in the money value of the transaction, consumption or sale in two different periods or points of time. Multiplying the quantity q by the price p of the commodity produced, transacted or sold gives the total money value pq of the production, transaction or sale. If instead of point of time, period of time is considered, a suitable average is to be taken and is to be expressed in the same units.

Suppose p_0 and q_0 denote the price and quantity of the commodity during the base period (0) and p_1 and q_1 denote the corresponding price and quantity during a given period (1). Then the total value of the commodity during the base period is $v_0 = p_0q_0$ and the corresponding total value during the given period is $v_1 = p_1q_1$. The value relative of the period 1 with respect to the base period 0 is defined by

Value relative in percentage (of period 1 with respect to 0)

$$= \frac{v_1}{v_0} \times 100 = \frac{p_1q_1}{p_0q_0} \times 100 \quad \ldots \ldots \ldots (7.3)$$

We denote value relative in percentage or without percentage by $v_{\frac{0}{1}}$.

Exercise 4: The production of tea in Assam in the month of January 2009 was 3608 tonnes and that in February 2009 was 3700 tonnes. The corresponding prices of tea were Rs. 100.00 and Rs. 120.00 per kilogram, respectively. Find the value relative.

Solution: Here the base period is January 2009 (say period J) and the given period is February 2009 (say period F). Then, q_J = 3608000 kg, p_J = Rs. 100.00 and q_F = 3700000 kg, p_F = Rs. 120.00.

Hence, $v_J = q_J \times p_J$ = Rs. 360800000.00

$v_F = q_F \times p_F$ = Rs. 444000000.00

Using formula (7.3), we find the required value relative,

$$v_{\frac{F}{J}} = \frac{444000000}{360800000} \times 100 = 123\%$$



CHECK YOUR PROGRESS

Q.1: Write True (T) or False (F):

i) Index numbers are specialized averages to measure relative changes. (T/F)

ii) Index numbers measure net change in related variables over time. (T/F)

iii) Simple index numbers require weights in their construction. (T/F)

iv) Index numbers do not indicate the trend of change in the economy. (T/F)

v) Price relative is defined as the ratio of current year's price to some reference year's price. (T/F)

Q.2: Fill in the blanks:

i) Index number is a statistical device to express ..... change in related variable.

ii) An index number is a ..... average.

iii) Index number measures net change in a group of ..... variables.

iv) In the index number represented by $P_{\frac{1990}{1998}}$, the base year is ..... and the current year is .....

v) Simple index numbers have ..... weights.

vi) The price relative means ratio of current price to .....

vii) If the price relative increases from 100 to 300, it shows ..... increase in prices.

13.5 PROPERTIES OF RELATIVES

Let p_a, p_b, p_c, ..., ..., ... denote the prices in the periods a, b, c, ..., ..., ... respectively. Also, let q_a, q_b, q_c, ..., ..., ... denote the quantities and v_a, v_b, v_c, ..., ..., ... denote the values for the periods respectively. The relatives satisfy some properties which are directly obtained from the definitions. We

shall state the results for price relatives and write similar results for the quantity and value relatives.

1) **Identity property:** The price relative of a given period with respect to the same period is 1, that is, $p_{\frac{a}{a}} = 1$.

2) **Time reversal property:** If the base period and the reference period are interchanged, then the product of the corresponding relatives is unity. That is, one is the reciprocal of the other. In other words,

$p_{\frac{a}{b}} = \frac{1}{p_{\frac{b}{a}}}$. We know that $p_{\frac{a}{b}} = \frac{p_b}{p_a} = \frac{1}{\frac{p_a}{p_b}} = \frac{1}{p_{\frac{b}{a}}}$

3) **Circular or Cyclic property:** This property states that,

$p_{\frac{a}{b}} \times p_{\frac{b}{c}} \times p_{\frac{c}{a}} = 1$. We can prove this property as follows:

We know that $p_{\frac{a}{b}} = \frac{p_b}{p_a}$, $p_{\frac{b}{c}} = \frac{p_c}{p_b}$, $p_{\frac{c}{a}} = \frac{p_a}{p_c}$

The required property follows by multiplying the above three terms.

4) **Modified circular or cyclic property:** This property states that,

$p_{\frac{a}{b}} \times p_{\frac{b}{c}} = p_{\frac{a}{c}}$. We prove this property as follows:

We have $p_{\frac{a}{b}} \times p_{\frac{b}{c}} = \frac{p_b}{p_a} \times \frac{p_c}{p_b} = \frac{p_c}{p_a} = p_{\frac{a}{c}}$

13.6 CONSTRUCTION OF SIMPLE INDEX NUMBERS

In general we are interested in general price level rather than on the price level of any particular item. So we have to consider a number of items at the same time. If each commodity is of different qualities then all such qualities have also to be taken into consideration. We have to consider simultaneously a number of items or commodities together for measuring change in price or production.

The sources of data must be reliable and care should be taken in selecting the sources of data. For example, official publications as well as other reliable unofficial reports from producers and merchants should be considered for computing index for production. If retail food prices are being

considered, then prices in stores, not in wholesale markets, are to be taken into consideration.

Now we shall discuss how to construct simple index numbers from a given set of data, like prices of certain commodities, production of certain commodities etc.

There are two methods of constructing simple index numbers. These are:

1) Simple aggregate method and
2) Method of simple average of relatives.

Remember that all items in a simple index number are assigned equal weights. These are unweighted index numbers.

1) Simple Aggregate Method: In this case each item is given equal weights. If we give an equal weight to each item, it means the same thing, whether each item is given a weight or not. It is the simplest method of constructing index numbers.

1.1) Simple aggregate method to find price index numbers: We use the following three steps to find price index number.

Step 1: Find the sum of the current year (that is, given period) prices of all the items included in the list. If 1 denote the current year, then find Σp_1.

Step 2: Find the sum of the base year prices of all the items included in the list. If 0 denote the base year, then find Σp_0.

Step 3: Use the formula $p_{\frac{0}{1}} = \frac{\sum p_1}{\sum p_0} \times 100$

Exercise 5: Find the price index number, using simple aggregate method, for the following information.

Commodities	Units	Prices (in Rupees)	
		2000	2005
Wheat	Quintal	800	1200
Rice	Quintal	1200	1600
Milk	Litres	12	18
Tea	Kilograms	100	120
Clothing	Metres	30	40
Meat	Kilograms	80	120

Solution: Represent prices of 2005 (current year) as p_1 and that of 2000 (base year) as p_0 and reconstruct the table as follows:

Commodities	p_0	p_1
Wheat	800	1200
Rice	1200	1600
Milk	12	18
Tea	100	120
Clothing	30	40
Meat	80	120
	$\Sigma p_0 = 2222$	$\Sigma p_1 = 3098$

Therefore, the price index of 2005 with base year 2000 is:

$$p_{\frac{0}{1}} = \frac{3098}{2222} \times 100 = 139.4$$

Thus, the increase in prices (or price level) is 139.40 – 100 = 39.4%.

Exercise 6: The prices of three commodities A, B and C increased from Rs. 50, Rs. 7 and Rs. 3 in 2005 to Rs. 52, Rs. 8 and Rs. 5 respectively in 2006. Using the simple aggregate method, find by how much on an average the prices have increased?

Solution: We know construct the following table:

Commodities	Price in 2005 (p_0)	Price in 2006 (p_1)
A	50	52
B	7	8
C	3	5
	$\Sigma p_0 = 60$	$\Sigma p_1 = 65$

Therefore, the price index of 2006 with respect to the base period 2005 is:

$$p_{\frac{2005}{2006}} = \frac{\sum p_1}{\sum p_0} \times 100 = \frac{65}{60} \times 100 = 108.33$$

This shows that the price index of current year 2006 with respect to base year 2005 is 108.33%. Hence the price increase in one year is 108.33 – 100 = 8.33%.

1.2) Simple aggregate method to find quantity index numbers:

We use the following three steps to find quantity index number.

Step 1: Find the sum of the current year (that is, given period) quantity or volume produced, consumed or transacted of all the items included in the list. If 1 denote the current year, then find Σq_1.

Step 2: Find the sum of the base year quantity or volume produced, consumed or transacted of all the items included in the list. If 0 denote the base year, then find Σq_0.

Step 3: Use the formula $q_{\frac{0}{1}} = \frac{\sum q_1}{\sum q_0} \times 100$

2) **Method of Simple Average of Relatives:** In this method, average of the relatives is obtained by using any one of the measures of central tendency. In particular, here we shall use arithmetic mean for averaging the relatives. We shall discuss briefly this method only for price relatives. Similar formulas on quantity index number and value index number can be obtained considering quantity and value relatives instead of price relative. There are few questions to find quantity and value index numbers in the section ‘‘Possible Questions’’. You should work out those problems with the help of the problems on price index numbers which are solved below.

2.1) Method of Simple Average of Price Relatives: We know that a price relative is nothing but the ratio of current year prices to those in the base year. Taking the arithmetic mean of the price relatives, we get the following formula for price index number:

$$p_{\frac{0}{1}} = \frac{\sum\left(\frac{p_1}{p_0} \times 100\right)}{N}$$

where N stands for number of commodities included in the index numbers.

2.2) Steps to calculate price index numbers by Simple Average of Price Relatives Method:

Step 1: Find percentage price relative for each commodity, that is,

$$\frac{p_1}{p_0} \times 100 .$$

Step 2: Find the sum of these percentage price relatives, that is,

$$\sum\left(\frac{p_1}{p_0} \times 100\right).$$

Step 3: Divide $\sum\left(\frac{p_1}{p_0} \times 100\right)$ by the total number of commodities included in the list.

Exercise 7: Give the solution of Exercise 5 using the Simple Average of Price Relatives.

Solution: We construct the following table:

Sl. No.	Commodities	Units	Prices (in Rupees)		Price Relative $\frac{p_1}{p_0} \times 100$
			2000 p_0	2005 p_1	
1	Wheat	Quintal	800	1200	$\frac{1200}{800} \times 100 = 150.00$
2	Rice	Quintal	1200	1600	$\frac{1600}{1200} \times 100 = 133.33$
3	Milk	Litres	12	18	$\frac{18}{12} \times 100 = 150.00$
4	Tea	Kilograms	100	120	$\frac{120}{100} \times 100 = 120.00$
5	Clothing	Metres	30	40	$\frac{40}{30} \times 100 = 133.33$
6	Meat	Kilograms	80	120	$\frac{120}{80} \times 100 = 150.00$
N = 6					$\sum\left(\frac{p_1}{p_0} \times 100\right) = 836.66$

Hence, the price index number of 2005 with respect to base year 2000 is:

$$P_{\frac{0}{1}} = \frac{\sum\left(\frac{p_1}{p_0}\times 100\right)}{N} = \frac{836.66}{6} = 139.44$$

Exercise 8: Give the solution of Exercise 6 using the Simple Average of Price Relatives.

Solution: We first prepare the following table:

Sl. No.	Commodities	Prices (in Rupees)		Price Relative
		2005 p_0	2006 p_1	$\frac{p_1}{p_0}\times 100$
1	A	50	52	$\frac{52}{50}\times 100 = 104.00$
2	B	7	8	$\frac{8}{7}\times 100 = 114.29$
3	C	3	5	$\frac{5}{3}\times 100 = 166.66$
N = 3				$\sum\left(\frac{p_1}{p_0}\times 100\right) = 384.95$

Hence, the price index number of 2006 with respect to base year 2005 is:

$$P_{\frac{0}{1}} = \frac{\sum\left(\frac{p_1}{p_0}\times 100\right)}{N} = \frac{384.95}{3} = 128.32$$

13.7 CONSTRUCTION OF WEIGHTED INDEX NUMBERS

In weighted index number each item is given weight according to the importance it occupies in the list. There are two groups of methods to calculate index number of this category: (a) Weighted aggregate method and (b) Weighted average of price relative methods. In this unit, we will study three types of weighted aggregate method. These methods are

popularly known as: (1) Laspeyre's Method, (2) Paasche's Method and (3) Fisher's Method. All the methods are used to calculate weighted index numbers. The main difference between the Laspeyre's Method and Paasche's Method is that Laspeyre's uses base year quantities of commodities as their relative weights, while Paasche's uses current year quantities of commodities as their relative weights for preparing a price index.

1) Laspeyre's Method:

Laspeyre's Price Index Number: It uses base year quantities as the weights. Accordingly, the formula is:

$$P_L = \frac{\sum p_1 q_0}{\sum p_0 q_0} \times 100$$

Laspeyre's Quantity Index Number: If we multiply the quantities by the base year price, then we get Laspeyre's quantity index number. Accordingly the formula for Laspeyre's quantity index number is:

$$Q_L = \frac{\sum q_1 p_0}{\sum q_0 p_0} \times 100$$

Steps to Calculate Weighted Index Number by Laspeyre's Method:

Step 1: Multiply current year price (p_1) with the base year quantity (q_0) to get p_1q_0 for each item/commodity.

Step 2: Multiply base year price (p_0) with the base year quantity to get p_0q_0 for each item/ commodity.

Step 3: Add all p_1q_0 and p_0q_0 separately to get Σp_1q_0 and Σp_0q_0.

Step 4: Divide Σp_1q_0 by Σp_0q_0 and multiply by 100 to obtain Laspeyre's price index number.

Exercise 9: Find Laspeyre's price index number from the following data.

Commodity	Price (in Rs.)		Quantity bought in units	
	1995	1996	1995	1996
A	50	52	10	12
B	7	9	3	4
C	3	5	7	6

Solution: Represent prices and quantities in current year by p_1 and q_1 respectively and represent prices and quantities in base year by p_0 and q_0 respectively. Let us construct the following table following the steps 1-3.

Commodities	p_0	q_0	p_1	p_0q_0	p_1q_0
A	50	10	52	500	520
B	7	3	9	21	27
C	3	7	5	21	35
				$\Sigma p_0q_0 = 542$	$\Sigma p_1q_0 = 582$

Therefore, the Laspeyre's price index number is:

$$P_L = \frac{582}{542} \times 100 = 107.38$$

2) Paasche's Method:

Paasche's Price Index Number: It uses current year quantities (q_1) as the weights. Accordingly, the formula is:

$$P_P = \frac{\sum p_1q_1}{\sum p_0q_1} \times 100$$

Paasche's Quantity Index Number: It uses current year prices (p_1) as the weights. Accordingly, the formula is:

$$Q_P = \frac{\sum q_1p_1}{\sum q_0p_1} \times 100$$

Exercise 10: Find Paasche's price index number from the data given in Exercise 9.

Solution: Represent prices and quantities in current year by p_1 and q_1 respectively and represent prices and quantities in base year by p_0 and q_0 respectively. Let us construct the following table.

Commodities	p_0	q_1	p_1	p_1q_1	p_1p_1
A	50	10	52	500	520
B	7	3	9	21	27
C	3	7	5	21	35
				$\Sigma p_0q_1 = 646$	$\Sigma p_1q_1 = 690$

Therefore the Paasche's price index number is:

$$P_P = \frac{690}{646} \times 100 = 106.8$$

Remark: You have noticed that the value of Laspeyre's index number (107.38) and the value of Paasche's index number (106.8) are different for the same data. The difference has arisen because of differences in weights.

3) **Fisher's Method:** The Fisher's price index number is nothing but the geometric mean of the Laspeyre's and Paasche's price index numbers. Hence the formula for Fisher's price index number is:

$$P_F = \sqrt{P_L \times P_p} = \sqrt{\left(\frac{\sum p_1 q_0}{\sum p_0 q_0} \times 100\right) \times \left(\frac{\sum p_1 q_1}{\sum p_0 q_1} \times 100\right)}$$

$$= \sqrt{\frac{\sum p_1 q_0}{\sum p_0 q_0} \times \frac{\sum p_1 q_1}{\sum p_0 q_1}} \times 100 \qquad \text{... ... ... (1)}$$

Similarly, the Fisher's quantity index number is given by:

$$Q_F = \sqrt{Q_L \times Q_P} = \sqrt{\left(\frac{\sum q_1 p_0}{\sum q_0 p_0} \times 100\right) \times \left(\frac{\sum q_1 p_1}{\sum q_0 p_1} \times 100\right)}$$

$$= \sqrt{\frac{\sum q_1 p_0}{\sum q_0 p_0} \times \frac{\sum q_1 p_1}{\sum q_0 p_1}} \times 100 \qquad \text{... ... ... (2)}$$

Exercise 11: Find Fisher's price index number from the data given in Exercise 9.

Solution: We have already found that for the given data,

$P_L = 107.38$ and $P_P = 106.8$

Hence the Fisher's price index number is:

$$P_F = \sqrt{107.38 \times 106.8} = 107.09$$



CHECK YOUR PROGRESS

Q.3: Write true (T) or False (F):

i) There are two types of index numbers: simple or unweighted and weighted index numbers. (T/F)

ii) There are two methods of constructing simple index numbers: simple aggregate method and simple average relatives. (T/F)

iii) We get the same index number whether we use simple aggregate method or simple average of price relatives' method. (T/F)

iv) Laspeyre's uses current year quantities as weights. (T/F)

v) Paasche's uses base year quantities as weights. (T/F)

Q.4: Fill up the blanks:

i) In Laspeyre's method of finding price index number we use ..... quantities as weights.

ii) In Paasche's method of finding price index number we use ..... quantities as weights.

iii) $I = \frac{\sum p_1 q_0}{\sum p_0 q_0} \times 100$ is ..... weighted index number.

iv) $I = \frac{\sum p_1 q_1}{\sum p_0 q_1} \times 100$ is ..... weighted index number.

Q.5: Calculate quantity index by (i) Laspeyre's method (ii) Paasche's method (iii) Fisher's method.

Commodity	2004		2006	
	Price (p_0)	Total value (p_0q_0)	Price (p_1)	Total value (p_1q_1)
A	10	100	12	144
B	12	144	14	196
C	14	196	16	256
D	16	256	18	324
E	18	324	20	400

Q.6: From the following date calculate Price index numbers for 2007 as base by (1) Laspeyre's method (2) Paasche's method and (3) Fisher's method.

Commodity	2006		2007	
	Price	Quantity	Price	Quantity
A	20	8	40	6
B	50	10	60	5
C	40	15	50	15
D	20	20	20	25

13.8 TEST OF ADEQUACY OF INDEX NUMBERS

We have seen in 13.6 section that the relatives have certain important properties. What is true for an individual commodity should also true for a group of commodities. The index numbers as an aggregative relative should also satisfy the same set of properties. We shall examine the properties that a good index number should have.

1) **Time Reversal Test or Property:** If the two periods, the base period and the reference period are interchanged, the product of the two index numbers should be unity. In other words, one should be the reciprocal of the other.

Denote the index number of the given year (1) relative to the base year (0) by $I_{\frac{0}{1}}$. Then $I_{\frac{1}{0}}$ is the index number of the year (0) relative to the year (1), that is, with the years interchanged time reversal test requires that,

$$I_{\frac{0}{1}} \times I_{\frac{1}{0}} = 1, \text{ or } I_{\frac{1}{0}} = \frac{1}{I_{\frac{0}{1}}}$$

An index number which satisfies the above property is said to satisfy time reversal test.

Remark: As we have seen in section 13.6, all the relatives satisfy the time reversal property.

Let us now examine whether Laspeyre's index number satisfy this property or not. We know that Laspeyre's index number (without

percentage) is $\frac{\sum p_1 q_0}{\sum p_0 q_0}$. Interchanging the time subscripts, we get

$\frac{\sum p_0 q_1}{\sum p_1 q_1}$ and hence their product is $\frac{\sum p_1 q_0}{\sum p_0 q_0} \times \frac{\sum p_0 q_1}{\sum p_1 q_1} \neq 1$.

Consequently, Laspeyre's price index number does not satisfy the time reversal property. Similarly, it can be seen that Laspeyre's quantity index number also does not satisfy the time reversal property. Also, Paasche's price and quantity index numbers do not satisfy this property.

Now consider the Fisher's price index number. Interchanging the time subscripts we find another index number. The product of the two is given by

$$\sqrt{\frac{\sum p_1 q_0}{\sum p_0 q_0} \times \frac{\sum p_1 q_1}{\sum p_0 q_1}} \times \sqrt{\frac{\sum p_0 q_1}{\sum p_1 q_1} \times \frac{\sum p_0 q_0}{\sum p_1 q_0}}$$

which is equal to 1. Hence Fisher's price index number satisfies the time reversal property.

2) **Factor Reversal Test or Property:** If the two factors *p* and *q* in a price index formula are interchanged, so that a quantity index number is obtained, then the product of the two index numbers should give the true value ratio, $\frac{\sum p_1 q_1}{\sum p_0 q_0}$

In other words, the price index number multiplied by the corresponding quantity index number should give the true ratio of value in the given year (1) to the value in the base year (0). This property holds for a single commodity, since

$$\frac{p_1}{p_0} \times \frac{q_1}{q_0} = \frac{p_1 q_1}{p_0 q_0}$$

but it does not hold for most of the index numbers. For example,

$$P_L \times Q_L = \frac{\sum p_1 q_1}{\sum p_0 q_0} \times \frac{\sum q_1 p_0}{\sum q_0 p_0} \neq \frac{p_1 q_1}{p_0 q_0}$$

Now consider the Fisher's index number. We have from the formulas (1) and (2),

$$P_F \times Q_F = \sqrt{\frac{\sum p_1q_0}{\sum p_0q_0} \times \frac{\sum p_1q_1}{\sum p_0q_1}} \times \sqrt{\frac{\sum q_1p_0}{\sum q_0p_0} \times \frac{\sum q_1p_1}{\sum q_0p_1}}$$

$$= \frac{\sum p_1q_1}{\sum p_0q_0}$$

This proves that Fisher's index number also satisfy factor reversal test. Fisher's index number is one of the rare index numbers which satisfy both the time reversal and factor reversal tests. So it is called "Fisher's Ideal Index Number".

3) **Circular Test:** This test is based on the shifting of the base period. If $I_{\frac{i}{j}}$ denotes the index number for the given period j with respect to the base period i, then this test requires that, $I_{\frac{a}{b}} \times I_{\frac{b}{c}} \times I_{\frac{c}{a}} = 1$

Remark: As we have seen in section 13.6, all the relatives satisfy the circular property.



13.9 LET US SUM UP

- Index number is a statistical measure or device with a purpose of showing average change in one or more related variables over time and space.
- Price index numbers are more commonly used. It measures relative changes in prices over a time period.
- We can have either simple or weighted index number. Simple index is also called unweighted index number or index number with equal weights. In this unit, we have discussed about the simple index numbers as well as weighted index number.
- We have discussed four properties of relatives: Identity property, time reversal property, circular property and modified circular property.

- We have learnt about two methods of constructing simple index numbers: Simple aggregate method and Method of simple average of relatives. The formula to find index number using simple aggregate method is given by $P_{\frac{0}{1}} = \frac{\sum p_1}{\sum p_0} \times 100$.

  Again, the formula to find price index number using the method of simple averages of relatives is given by $P_{\frac{0}{1}} = \frac{\sum p_1}{\sum p_0} \times 100$.

- We have also leant about weighted index numbers. There are two groups of methods to calculate index number of this category: (a) Weighted aggregate method and (b) Weighted average of price relative methods. We briefly discussed about three types of weighted aggregate method. These methods are popularly known as (1) Laspeyre's Method, (2) Paasche's Method and (3) Fisher's Method.
- We have also learnt about the test of adequacy of index numbers. We mentioned three properties that a good index number should have. Fisher's index number satisfies both the time reversal and factor reversal tests and hence it is known as Fisher's Ideal Index Number.



13.10 FURTHER READING

1) Agarwal, D. R. (2006). *Business Statistics*. Delhi: Vrinda Publications.
2) Gupta S. C. (1994). *Fundamentals of Statistics*. New Delhi: Himalayan Publishing House.
3) Rajagopalan, S. P. & Sattanthan, R. (2009). *Business Statistics and Operations Research*. New Delhi: Tata McGraw-Hill
4) Sharma, J. K. (2007). *Business Statistics*. New Delhi: Pearson Education Ltd.
5) Verma, A. P. (2007). *Business Statistics*. Guwahat: Asian Books Private Limited.

13.11 ANSWERS TO CHECK YOUR PROGRESS

Ans. to Q. No. 1: i) T, ii) T, iii) T, iv) F, v) F, vi) T

Ans. to Q. No. 2: i) average, ii) specialized, iii) related, iv) 1990, 1998, v) equal, vi) base year price, vii) 200%

Ans. to Q. No. 3: i) T, ii) T, iii) F, iv) F, v) F

Ans. to Q. No. 4: i) base year, ii) current year, iii) Laspeyre's, iv) Paache's

Ans. to Q. No. 5: Construction of various Indicer

p_0	q_0	p_1	q_1	p_0q_0	p_0q_1	p_1q_1	p_1q_0
10	10	12	12	100	120	144	120
12	12	14	14	144	168	196	168
14	14	16	16	196	224	256	224
16	16	18	18	256	288	324	288
18	18	20	20	324	360	400	360
			Total	1020	1160	1320	1160

i) Laspeyr's Price Index is: $\frac{\sum p_1q_0}{\sum p_0q_0} \times 100$

$= \frac{1160}{1020} \times 100$

$= 113.7$

ii) Paasche's Price Index is: $\frac{\sum p_1q_1}{\sum p_0q_1} \times 100$

$= \frac{1320}{1160} \times 100$

$= 113.8$

iii) Fisher's Price Index is: $\sqrt{\frac{\sum p_1q_0}{\sum p_0q_0} \times \frac{\sum p_1q_1}{\sum p_0q_1}} \times 100$

$= \sqrt{113.7 \times 113.8}$

$= 113.75$

Ans. to Q. No. 6: Computation of Index Numbers

Commodity	2006		2007					
	p_0	q_0	p_1	q_1	p_1q_0	p_0q_0	p_1q_1	p_0q_1
A	20	8	40	6	320	160	240	120
B	50	10	60	5	600	500	300	250
C	40	15	50	15	750	600	750	600
D	20	20	20	25	400	400	500	500
					Σp_1q_0 2070	Σp_0q_0 1660	Σp_1q_1 1790	Σp_0q_1 1470

i) Laspayer's Index: $\frac{\sum p_1q_0}{\sum p_0q_0} \times 100$

$= \frac{2070}{1660} \times 100$

$= 124.70$

ii) Paasche's Index: $\frac{\sum p_1q_1}{\sum p_0q_1} \times 100$

$= \frac{1790}{1470} \times 100$

$= 121.77$

iii) Fisher's Index: $\sqrt{\frac{\sum p_1q_0}{\sum p_0q_0} \times \frac{\sum p_1q_1}{\sum p_0q_1}} \times 100$

$= \sqrt{\frac{2070}{1660} \times \frac{1790}{1470}} \times 100$

$= \sqrt{1.5184} \times 100$

$= 1.2322 \times 100$

$= 123.22$



13.12 MODEL QUESTIONS

Q.1: Explain the meaning of an index number and its main characteristics.

Q.2: "Index numbers are economic barometers." Why are index numbers so called? Explain.

Q.3: What is price relative? Show with the help of an example how it is calculated?

Q.4: Define quantity and value relatives. What do they measure?

Q.5: State the important properties of relatives.

Q.6: Explain the method of construction of unweighted index numbers by Simple Aggregate Method.

Q.7: Explain the method of construction of unweighted index numbers by Simple Aggregate of Price Relative Method.

Q.8: The exchange rate of a US dollar was Rs. 44.00 in July 2008 and was Rs. 50.00 in February 2009. Find the price relative.

Q.9: The production of rice in Assam in the month of January 2008 was 3608 tonnes and that in February 2009 was 3700 tonnes. Find the quantity relative.

Q.10: The following table shows the prices of the commodities consumed in 1970 and 1978.

		1970	1978
Commodities	Unit	Price (Rs.)	Price (Rs.)
A	Kilogram	4.00	4.60
B	Litre	8.00	8.75
C	Metre	7.00	8.00
D	Kilogram	14.00	16.00
E	Litre	3.00	4.00

Find the price index number using:

a) simple aggregate method

b) method of simple aggregate of price relatives.

Q.11: The production of rice, milk, cloth, wheat and petrol are given in the following table. Taking 1980 as base year, compute quantity index number using:

a) simple aggregate method

b) method of simple aggregate of quantity relatives.

		Quantity Produced	
Commodities	Unit	1980	1990
Rice	Kilogram	400000	700000
Milk	Litre	800000	905000
Cloth	Metre	100000	105000
Wheat	Kilogram	140000	160000
Petrol	Litre	300500	400000

Q.12: What do you mean by tests of index numbers?

Q.13: Explain the important criterion or test that a good index number should satisfy.

Q.14: The prices and quantity of five commodities are given in the following table. Taking 2000 as base year, compute value index number using:

a) simple aggregate method

b) method of simple aggregate of value relatives.

Commodities	2000		2002	
	Price (Rs.)	Quantity (Kg.)	Price (Rs.)	Quantity (Kg.)
A	400	10	700	12
B	800	15	1905	17
C	100	5	125	5
D	140	7	160	7
E	300	6	400	6

Q.15: Find Laspeyre's and Paasche's price and quantity index number from the following data. Also find the Fisher's price and quantity index numbers.

Commodities	Base year		Current year	
	Price (Rs.) p_0	Quantity (Kg.) q_0	Price (Rs.) p_1	Quantity (Kg.) q_1
A	5	25	6	30
B	10	5	15	4
C	3	40	2	50
D	6	30	8	35

Q.16: The following table shows the price quantity of the commodities consumed in 2000 and 2006.

Commodities	2000		2006	
	Price (Rs.)	Quantity (Kg.)	Price (Rs.)	Quantity (Kg.)
I	400	35	460	35
II	800	10	1000	10
III	300	15	50	12
IV	4000	10	5000	8
V	50	20	300	20

Find the following index numbers.

1) Laspeyre's Price and Quantity index numbers.
2) Paasche's Price and Quantity index numbers.
3) Fisher's Price and Quantity index numbers.

*** ***** ***

UNIT 14 : TIME SERIES

UNIT STRUCTURE

14.1 Learning Objectives

14.2 Introduction

14.3 Definition of Time Series

14.4 Importance of Time Series Analysis

14.5 Components of a Time Series

14.6 Methods of Measuring Secular Trend

14.7 Estimation of the Trend by the Method of Moving Average

14.8 Let Us Sum Up

14.9 Further Reading

14.10 Answers To Check Your Progress

14.11 Model Questions

14.1 LEARNING OBJECTIVES

After going through this unit, you will able to:

- understand the concept of time series
- explain the importance of time series analysis
- known about the various factors which affect a time series
- know about the methods for determining secular trend in time series.

14.2 INTRODUCTION

In Economics, Business and Commerce, it is important to make estimate for the future e.g. an economist is interested to know the figures of national income, prices and wages, population etc. for his future planning or a businesman likes to estimate his likely sales in the coming years to adjust his production accordingly. For making such estimates, one has to collect information from the past i.e. one has to deal with statistical data collected and recorded at successive intervals of time (or points of time). Such statistical data relating to time all refered to as Time Series. The analysis of time series has an important role in Economics, Business and

Commerce. Basically, most of the statistical technigues for the analysis of time series data have been developed by economists. However, these techniques can also be applied for study of behaviour of any phenomenon collected chronologically over a period of time in any discipline relating to natural and social sciences, though not directly related to business and economics.

14.3 DEFINITION OF TIME SERIES

A Time Series is a set of observations taken at specified times, usually (but not always) at equal intervals. Thus a set of data depending on time (which may be year, half-year, quarter, month, week, days etc.) is called a time series.

Some examples of time series are:

i) The annual production of steed in India over the last ten year.

ii) The sale of a departmental store in different months of the year.

iii) The daily closing price of a share in the Bombay Stock exchange for a month.

iv) Hourly Temperature recorded by the meteorological office in a city.

v) Quarterly Consumer Price Index numbers of a country for last three years.

vi) Daily rainfall of a particular city in the month of May, June, July and August.

Mathematically, a time series is defined by the functional relationship:

$$Y = f(t)$$

Where y is the value of the phenomenon (or variable) under consideration at time t. Thus, if the values of a phenomenan (variable) at times $t_1, t_2, ..., ..., ..., t_n$ are $y_1, y_2, ..., ..., ..., y_n$ respectively, then the series

t:	t_1	t_2	t_3	...,	...,	...,	t_n
y:	y_1	y_2	y_3	...,	...,	...,	y_n

Constitutes a time series.

14.4 IMPORTANCE OF TIME SERIES ANALYSIS

Analysis of time series is of special importance to Businessman, Economist, Scientist, Sociologist, Geologists and Research workers in various discipline. The following points indicate the importance of time series analysis:

1) **Past Behaviour:** It helps us to understand the past behaviour of a variable. Analysis of past data discloses effect of various factors on the variable under study.
2) **Forecasting:** With the knowledge of past behaviour, it would be possible, within certain limits, to forecast the future behaviour of the variable under study. This helps in making future plans of action. Various five year plans of our country are based on the analysis of past data.
3) **Evaluation:** It halps in evaluation of current achievements. The review and evaluation of progress made on the basis of a plan are done on the basis of time series data. For example, the progress of various plans in our country is judged by the yearly rate of growth in GNP (Gross National Product).
4) **Comparison:** Analysis of time series helps in making comparison between one time period to another.It also facilitate to compare the actual values with the expected values. It provides a scientific basis for making comparisons by studying and isolating the causes of variation of the variable under study.

14.5 COMPONENTS OF A TIME SERIES

Analysis of time series shows that if the values of a phenomenon are observed at different periods of time, the values so obtained are always fluctuating. The reason behing this type fluctuation is the joint effect of various factors (or forces). For example, the sales (y) of a product depends on: (i) advertisement expenditure (ii) the price of the product (iii) the income of the consumer (iv) other competitive product in the market (v) tastes, fashions, habits and customs of the people and so on similarly, the price of a particular

product depends on its (i) demands (ii) various competitive products in the market (iii) availability of raw matirials (iv) transportation cost etc.

To know the proper nature of a time series, the statisticians have broadly classified the various operating factors (forces) affecting the values of a shenomenon in a time series into four parts. These are commonly known as **Components of a time series.** The four components of a time series are:

a) Secular Trend or Trend (T)
b) Seasonal Variation (S)
c) Cyclical Variation (C)
d) Irregular or Random Variation (I)

N.B.: Instead of the word 'variation', the word 'fluctuation' or 'movement' may also be used.

The changes in time series data are the result of the combined effect of these four components. Now, we proceed to a brief discussion on the various components of time series.

a) Secular Trend or Trend (T): The general tendency of the time series data showing continious growth (increase), stagnation or decline during a long period of time is called the **Secular Trend** or simply **Trend.** This phenomenon is usually observed in most of the time series relating to Business and Economics. e.g. an upward tendency is usually observed in time series relating to population, prices, income, production and sales of products, money in circulation etc, while a downward tendency is noticed in time series relating to birth rate, death rate, death rate due to epidemics etc due to the advancement in medical science, improved in medical facilities, better sanitation diet etc. Simpson and Kafka defined secular trend as: **"Trend, also called Secular Trend or Long term trend, is the basic tendency of a time series to grow or decline over a period of time. The concept of trend does not include short-range oscillations, but rather the steady movement over a long time."**

Remarks: It should be understood that the term 'long period' is a relative term and can not be defined exactly. In certain phenomenon,

a period of few hours may be long enough whereas in some others even a period as long as 3-4 years may not be regarded as a long period. For example, if some agricultural or industrial production of a partcular product shows an increase over the past 3 years, this can not be termed as a long term tendency, for which we must have data for 7-8 years. On the other hand, while a patient is having high fever and as per advice of the attending physician, temperature reading are needed to be recorded on every 4 hours, in order to observe whether the effect of a particular medicine shows a decline in temperature or not, then in such a case, taking reading for a period of 48 hrs to 72 hrs would be termed as a long period.

Uses of Trend: The study of the data over a long period of time enables us to have a general idea about the pattern of the behaviour of the phenomenon under consideration. This helps in business forecasting and planning of the future operations. Trend analysis enables us to compase two or more time series over different periods of time and draw important conclusions about them.

b) **Seasonal Variations (S):** The fluctuations or variations in the values of a time series over a period of one year or less are termed as seasonal variation. This variation occurs due to forces which are rhythmic in nature and which repeat themselves periodically over a span of less than a year. Such type of variations are present in a time series if the data are recorded quarterly (after every three months), monthly, weekly or daily basis. **Thus in a time series data where only annual figures are given, there can not be seasonal variation.** Most of the economic time series are influenced by swings e.s. prices, production and consumption of commodities, sales and profits of departmental stores, bank clearings and bank deposits etc. are all affected by seasonal variation. Seasonal variation takes place due to the following two causes–

   i) **Narural Causes:** As the name suggests, the various seasons or weather conditions and climatic changes play an important role in seasonal variation. For example, the sale of umbrella

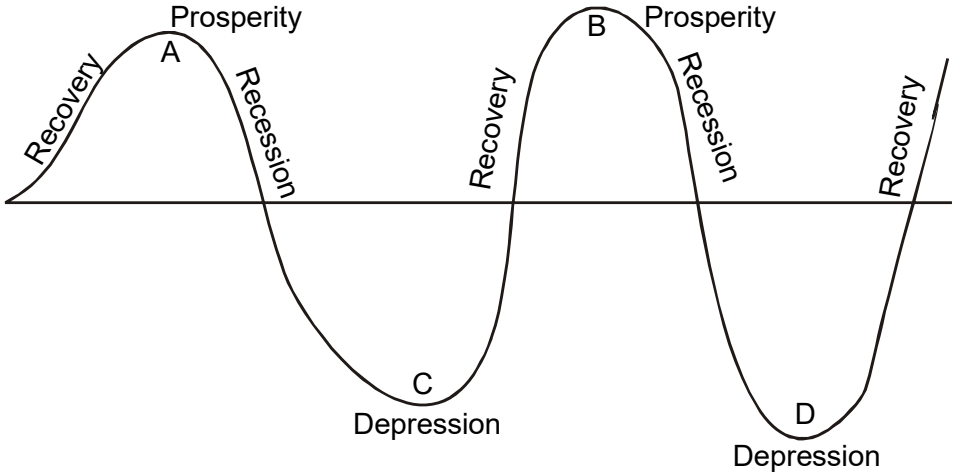
picks up very fast in raing season, the sale of woollen garment is more in winter, during summer demand for cold drinks and ice-cream increases. The prices of agricultural commodities always go down at the time of harvest and then pick up gradually.

ii) **Rituals and Social Customs:** Man made rituals, social customs and teaditions are also responsible for seasonal fluctuation of a time series. For example, sale of jewellary goes up during diwali and marriage season, in the beginning of an academic session sale of books, papers, uniforms etc. go up, on the eve if new year, sale of greetings card increases.

Uses of Seasonal Variation: The main objective of seasonal variations is to isolate them from trend and study their effects. A study of the seasonal pattern is extremely useful for a businesman, producers sales managers etc. in planing future operations and in formulation of policy decisions regarding purchase, productions, inventory control and advertising programmes. In the absence of any knowledge of seasonal variation, a seasonal upswing may be mistaken as indicator of better business conditions while a seasonal slump may be misinterpreted as deteriorating business conditions.

c) **Cyclical Variation (C):** Cyclical variations are the oscillatory mavement in a time series with period of oscillation greater than one year. The cyclical vanations though more or less regular, are not necessairly uniformly periodic i.e. they may or may not follow exactly similar patterns after equal intervals in time one cyclic period normally lasts from 7 to 9 years.

Cyclical variations are found to exist in almost all busimess and economic time series where it is known as **business cycle** or **trade cycle**. The ups and downs (or rises and decline) in business recurring at intervals of time are the effects of cyclical variation. A **business cycle consists of four phasses** namaly **prosperity** (boom), **recession, depression** and **recovery.** Each phase changes gradually into the next phase in the given order until one business cycle is completed:



(Business Cycle)

The time period between two successive booms or depressions (e.g. time period between A and B or C and D) is known as the length of a cycle.

Uses of Cyclical Variations: The study of cyclical variations is of great importance to business executives in the formulation of policies animed at stabilising the level of business activity. A knowledge of cyclical variations enables a businessman to have an idea about the periodicity of booms and depressions and accordingly timely steps may be taken for maintaing stable market for his product.

Note: Though seasonal variations and syclical variations are both periodic in nature, there is a significant difference between the two types of variations. These are mentioned below:

	Seasonal Variations		**Cyclical Variations**
i)	Seasonal variation takes place within one yeas.	i)	Cyclical variation takes place in a period more than one year. It voually takes place in 3 to 10 years time period.
ii)	In seasonal variation the periodic time remains the same.	ii)	In cyclical variation, the periodic time does not remain the same. It takes place only with some rough regularity.

iii)	It is mainly attributed to climatic changes and man made rituals and social customs.	iii)	Economic factors like price, production, sales, demand etc. are responsible for cyclical variations.

d) Irregular Variations (I): Irregular or Random Variations are such variations inherent in a time series which are caused by factors irregular (or erratic) in nature. These are purely random and unpredictable. These include all movements not already covered in trend, seasonal variation and cyclical variation. Irregular variations are caused by unforeseen factors like flood, famines, wars, earth. quakes, fire, epidemics, strike of trade union, terrorision etc. As there are unpredictable, they can not be controlled by human hand. For almost all the time series, irregular variations are inevitable. Normally, these are short term fluctuation but sometimes their effect is so intense that they may give rise to new cyclical or other movements.

Because of their absolutely random character, it is not possible to study them exclusively, nor we can forecast or estimate these precisely. The best that can be done about such variations is to obtain their rough estimates (from the past experience) and accordingly make provisions for such abnormalities during normal time in business.

14.6 METHODS OF MEASURING SECULAR TREND

There are four methods which are generally used for the study and measurement of secular trend in a time series. There are:

i) Graphical Method
ii) Semi-Average Method
iii) Moving Average Method
iv) Method of Leaot Squars.

Note: Here we will restrict our discussion to the **Method of moving average** only.

14.7 ESTIMATION OF TREND BY THE METHOD OF MOVING AVERAGE

The method of moving average **is a very simple and flexible method of measuring trend. This method consists in taking averages (arithmetic mean) of the values for a certain time span and placing it at the centre of the time span.**

The moving average is characterioed by a constant known as **period** of the moving average. Thus the moving average of period 'm' is a series of successive averages (Arithmetic means) of 'm' overlapping values at a time starting with 1st, 2nd, 3rd value and so on. Thus, for the time series values y_1, y_2, y_3, y_4, y_5, ..., ..., ..., the moving average (M.A.) values of period 'm' are given by:

$$\text{1st M.A.} = \frac{1}{m}(y_1 + y_2 + ... + y_m)$$

$$\text{2nd M.A.} = \frac{1}{m}(y_2 + y_3 + ... + y_{m+1})$$

$$\text{3rd M.A.} = \frac{1}{m}(y_3 + y_4 + ... + y_{m+2})$$

and so on we shall discuss two cases:

Case (i) **When period is odd:** It the period 'm' of the moving average is odd, then the successive values of the moving averages are placed against the middle most observation of the corresponding time period. For example, if m = 5, the first moving average value is placed against the 3rd value of the series, the second moving average value is placed against the 4th value of the time series and so on.

Case (ii) **When period is even:** It the period 'm' of the moving average is even, then there are two middle periods and the moving average values are placed in between the two middle periods of the time interval it covers. Obviously, in this case the moving average values will not coincide with a period of the given time series and an attempt is made to synchronise them with the original data by taking a two-

period average of the moving averages and placing them in between the corresponding time periods. This technique is called **centering** and the corresponding moving average values are called **centred moving average.** For example, if m = 4, the first moving average value is placed against the middle of 2nd and 3rd time intervals, the second moving average value is placed in between 3rd and 4th time period and so on. These values are given by–

$$\overline{y_1} = \frac{1}{4}(y_1 + y_2 + y_3 + y_4)$$

$$\overline{y_2} = \frac{1}{4}(y_2 + y_3 + y_4 + y_5)$$

$$\overline{y_3} = \frac{1}{4}(y_3 + y_4 + y_5 + y_6) \text{ and}$$

The centered moving averages are obtained on taking 2 period moving average of $\overline{y_1}$, $\overline{y_2}$, $\overline{y_3}$ and so on.

Thus, First centered moving average = $\frac{1}{2}\left(\overline{y_1} + \overline{y_2}\right)$

Second centered moving average = $\frac{1}{2}\left(\overline{y_2} + \overline{y_3}\right)$ and so on.

These centered moving averages are placed against 3rd, 4th, 5th period and so on.

MERITS AND DEMERITS OF MOVING AVERAGE METHOD:

Merits:

i) This method does not involve any mathematical complexities and is quite simple to understand and use.

ii) There is some inherent flexibility in this method. Addition of few more observations to the existing data does nt affect the trend values already obtained. This will simple result in some more values at the end.

iii) If the period of moving averages coincides with the period of cyclic fluctuations in the data, then this method completely eliminalis the effect of cyclical fluctuations.

iv) This method is also used for measurement of seasonal, cyclical and irregular variations.

Demerits:

i) One of the prime limitations of moving average method is that we can not obtain the trend values for all the given observations. Some trend values in the beginning and at the end have to forego depending on the period of moving average.

ii) This method is not suitable for the purpose of forecasting.

iii) The selection of period of moving average is a difficult task. Therefore, great care has to be taken in selecting the period, particularly, when there is no business cycle is the time series.

iv) If the time series reveals lineartrend only then this method is applicable.

Example 1: Calculate 3 yearly moving average from the data given below:

Year:	1980	1981	1982	1983	1984
Production (’oootons):	112	117	120	123	129
Year :	1985	1986	1987	1988	1989
Production (’oootons):	134	140	146	149	152

Solution: Table for calculation of 3 yearly moving averages.

Year	Prodiction (’ooo tons)	3 Yearly Moving Total	3 Yearly Moving Average
(1)	(2)	(3)	(4)
1980	112		
1981	117	349	116.3
1982	120	360	120
1983	123	372	124
1984	129	386	128.7
1985	134	403	134.3
1986	140	420	140
1987	146	435	14.5
1988	149	447	149
1989	152		

Explanation: In column (3) total of the first three entries namely 112, 117 and 120 is written against the middle value i.e. 117 of this set. Then total of next. three entries i.e. 117, 120 and 123 is written

against 120 and so on. Figures of column (4) are obtained by dividing the columns of figures (3) by 3.

Note: These 3 yearly moving average values are the trend values for the given time series.

Example 2: Calculate trend values by the method of 4 yearly moving average from the following data.

Year:	1991	1992	1993	1994	1995	1996
Value:	44	64	59	52	57	71
Year:	1997	1998	1999	2000	2001	2002
Value:	66	64	71	77	82	80

Solution: Table for calculation of 3 yearly mouing averages

Year	Value moving	4 yearly moving Total	4 yearly moving average	2-period average column	Centred (trend value)
(1)	(2)	(3)	(4)	(5)	(6) = (5) ÷ 2
1991	44				
1992	64 219	54.75			
1993	59 232	58		112.75	56.375
1994	52 239	59.75		117.75	58.875
1995	57 246	61.5		121.25	60.625
1996	71 272	68		126	63
1997	66 272	69.5		132.5	68.75
1998	64 278	69.5		137.5	68.75
1999	71 294	73.5		143	71.5
2000	77 310	77.5		151	75.5
2001	82				
2002	80				

Explanation: In column (3) total of the first four entries namely 44, 64, 59 and 52 is written against the middle of this set i.e. in between 64 and 59. Then the total of next four entries i.e. 64, 59, 52 and 57 is written in between 59 and 52 and so on.

Figures of column (4) are obtained by dividing the figures of column (3) by 4.

In column (5) the total of the first two entries of column (4) i.e. 54.76 and 58 is put in between them i.e. corresponding to the year 1993. Then the total of second and third entries of column (4) i.e. 58 and 59.75 is put corresponding to the year 1994 and so on.

The figures of the column (6) are obtained by dividing the entries of column (5) by 2.

Note: The 4 yearly centred moving averages are the trend values for the given time series.

Example 3: From the data given below, determine trend values by 5 yearly moving average method.

Solution:

Year:	1970	1971	1972	1973	1974	1975	1976	1977	1978
Sales (Rs. lakhs)	130	127	124	135	140	132	129	127	145

Table for calculation trend values by 5 yearly moving average method

Year	Sales (Rs. Lakhs)	5 Yearly Moving Total	5 Yearly Moving Average (Trend value)
(1)	(2)	(3)	(4)
1970	130		
1971	127		
1972	124	656	131.2
1973	135	658	131.6
1974	140	660	132.0
1975	132	663	132.6
1976	129	673	
1977	127		
1978	145		

Note: The first moving total 656 of column (3) is the sum of first five values 14, 130, 127, 124, 135 and 140. The second moving total is 127 + 124 + 135 + 140 + 132 = 658 which can also be obtained by adding (132 – 130) = 2 with first moving total 656. Similarly, 3rd moving Total is 658 + (129 – 127) = 660 and so on.



CHECK YOUR PROGRESS

Q.1: Fill in the blanks:

i) A time series consists of data arranged .....

ii) An overall rise or fall in the time series is called .....

iii) ..... can be made with the help of time series.

iv) Cyclical variations are caused by .....

v) ..... is the overall tendency of the time series data ..... or ..... over a ..... period of time.

vi) Short term variations are classified as (a) ..... (b) .....

vii) The four phases of business cycle (in order) are ....., ..., ...

viii) For the annual data ..... component is absent.

ix) The most important factors causing seasonal variations are .....

Q.2: With which component of time series would you associate each of the following.

i) An era of prosperity.

ii) An after puja sale in a departmental store.

iii) A strike in a factory delaying production by 4 weeks.

iv) The sale of umberlla in rainy season.

v) Secrease in death rate due to advancement of medical science.

vi) A need for increased rice production due to constant increase in population.

Q.3: Select the correct alternatives out of the given one–

a) A time series is a set of data recorded:

i) at time intervals ii) at successive point of time

iii) periodically iv) All the above

b) The decline in birth rate is attached to:

i) Secular Trend ii) Irregular Variation

iii) Seasonal Variation iv) Cyclical Variation

c) The sale of cold drink in summer season is attached to:

i) Secular Trend ii) Seasonal Variation

iii) Cyclical Variation iv) Irregular Variation

d) Seasonal variation repeats during a period of:

i) 5 years ii) 4 years

iii) 7 years iv) None of the above

e) Salient factors responsible for seasonal variation are:

i) Weather ii) Festival

iii) Social Custom iv) All the alove

f) Increase in death rate due to an epidemic is attached to:

i) Secular Trend ii) Cyclical Variation

iii) Seasonal Variation iv) None of the alove



14.8 LET US SUM UP

In this unit we have learnt about the following:

- Time series is a set of observation recorded at successive periods of time.
- The factors which affect the components of time series.
- The component of time series responsible for long term movement in a time series is called secular trend or simply trend.
- The component of time series that shows a regular and periodic pattern of movement over a time span of less than 1 year is called seasonal component.

- Cyclical variation is an oscillatory movement in a time series with period of oscillation greater than 1 year.
- Irregular movement are such variations inherent in a time series.
- Analysis of time series helps us to understand past behaviour, forecasting, evaluation of current achievement and making comparison between two time period.
- There are four methods for studying and measuring secular trend.
- One of the methods of measuring trend is 'Moving average method' which involve smoothing a time series by averaging succesive groups of data for a certain time span.



14.9 FURTHER READING

1) Agarwal D. R. *Business Statistics*. Delhi: Vrinda Publication.
2) Arora, P. N.; Arora, Sumeet; Arora, S. *Comprehensive Statistical Methods*. S. Chand & Company.
3) Ghosh, R. K. and Saha, S. *Business Mathematics and Statistics.* New Central Book Agency (P) Ltd.
4) Gupta, S. C. *Fundamentals of Statistics*. New Delhi: Himalaya Publishing House.
5) Gupta S. C and Gupta Indira. *Business Statistics.* Himalaya Publishing House.
6) Sharma J. K. *Business Statistics.* New Delhi: Pearsor Education Ltd.



14.10 ANSWERS TO CHECK YOUR PROGRESS

Ans. to Q. No. 1: i) Chronologically, ii) Trend / Secular Trend, iii) Forecasts, iv) Business / Trade Cycles, v) Secular Trend, increase, decrease, long, vi) (a) Seasonal (b) Cyclical (vii) Boom, Recession, Depression, Recovery, viii) Seasonal, ix) Weather and Social Customs.

Ans. to Q. No. 2: i) Cyclical, ii) Seasonal, iii) Irregular, iv) Seasonal, v) Secular, vi) Secular.

Ans. to Q. No. 3: a) (iv), b) (i), c) (ii), d) (iv), e) (iv), f) (iv)



14.11 MODEL QUESTIONS

Q.1: Define time series.

Q.2: Explain various components of time series.

Q.3: What is meant by Time Series Analysis? Discuss its importance in business.

Q.4: What do yoy meand by Secular Trend? Explain by giving examples.

Q.5: Explain the meaning of seasonal variations, with illustrations.

Q.6: What are cyclical variation? How are they caused?

Q.7: How do cyclical variations differ from seasonal variation.

Q.8: Which component of time series is associated with the following:

i) Recession

ii) A fire in a factory delaying production for three weeks.

iii) An era of depression.

iv) Sale of ice-cream in summer.

v) Increase in the production of rice in Assam due to application of modern technique.

Q.9: Obtain Trend Value by 5 yearly moving average method.

Year:	1990	1991	1992	1993	1994	1995	1996	1997	1998
Value:	242	250	252	249	253	255	251	257	260
Year:	1999	2000							
Value:	265	262							

Q.10: Calucate Trend Values by 3 yearly moving average method.

Year:	1981	1982	1983	1984	1985	1986	1987	1988
Sugar Production (lakh tons):	37.4	31.2	38.7	39.6	47.6	42.6	48.5	56.2
Year:	1989	1990						
Sugar Production (lakh tons):	60.6	58.2						

Q.11: Calculate trend values by 4 yearly moving average method.

Year:	1985	1986	1987	1988	1989	1990	1991	1992	1993
Value:	39	59	53	46	51	65	59	58	65
Year:	1994	1995	1996						
Value:	71	76	74						

Q.12: Determine the period of the moving average for the data given below and calculate moving average value for that period.

Year:	1	2	3	4	5	6	7	8	9
Value:	130	127	124	135	140	132	129	127	145
Year:	10	11	12	13	14	15			
Value:	158	153	146	145	164	170.			

[**Hint :** Since the peaks of the given data occur at the years 1, 5, 10 and 15, the data clearly exhibits a regular cyclical movement with period 5, hence the period of moving average for determining trend valus is also 5 i.e. period of cyclical variations]

*** ***** ***

UNIT 15: BASIC CONCEPT IN SAMPLING

UNIT STRUCTURE

15.1 Learning Objectives

15.2 Introduction

15.3 Defining Population, Parameter and Sampling

15.3.1 Random and Non-Random Sampling Methods

15.3.2 Sampling with and without Replacement

15.3.3 Large Samples

15.4 Errors in Sampling Procedure

15.5 Let Us Sum Up

15.6 Further Reading

15.7 Answer to Check Your Progress

15.8 Model Questions

15.1 LEARNING OBJECTIVES

After going through this unit, you will be able to:

- discuss the concept of a parameter and a statistic
- explain the concepts of a sampling with and without replacement
- explain random and non-random sampling
- discuss about standard error
- discuss about different methods of sampling
- explain merits and limitations of sampling.

15.2 INTRODUCTION

Often we are interested in drawing valid conclusions about a large group of individuals or objects. Instead of examining the entire group, which may be difficult to do, we examine only a small part. We do this with the aim of inferring certain facts about the group of individuals.

The entire group of individuals is known as population. The word ‘population’ used in statistics has altogether a different meaning.

You all know the dictionary meaning of the word population. But in mathematics we use the word with a bit different meaning. Mathematician use the word population to refer not only to people but to all items that have been chosen for Study. When it is possible to examine every item in the population that we have taken for study, then it is called a Census. But complete enumeration is not always possible in all situations. Then we take another procedure, which we will refer as sampling method. Mathematician use the word sample to describe portion chosen from the population. For example we have population of heights of all female inhabitants of Guwahati or a population of incomes of families living in Guwahati etc. A population may be finite, infinite or hypothetical (for example, outcomes of tossing a coin infinite number of times).

Often you have seen in news paper that Government estimated the production of rice in a year. Do they take complete enumeration? Certainly not. When they do this they select some areas and measure the production there and after that they do some mathematical calculation. This is just an example. Every estimation is done like this. Care should be taken in selecting the areas; otherwise you will not get good result. Sample must be the representative of population.

15.3 DEFINING POPULATION, PARAMETER AND SAMPLING

In mathematics, we describe samples and population by using measures such as mean, median, mode and standard deviation (s.d). When these terms describe the characteristics of a sample then they are called ***Statistics***. A 'Statistic' is a function of the sample. Examples are arithmetic mean $\left(\bar{x}\right)$ and sample variances (σ^2).

When they describe the characteristics of a population they are called ***Parameters***.

A ***population*** is the totality of any kind of units under consideration. It may be finite or infinite.

A sample is any portion of the population selected for study. The process of obtaining samples is called ***sampling***.

The fundamental assumption underlying the theory of sampling is random sampling, which consists in selecting the individuals from the populations, in such a way that each individuals of the population has the equal chance of being selected.

Two types of sampling will be discussed– (1) Random Sampling and (2) Non random Sampling.

Purpose of Sampling: A sample is not studied for its own sake. The main objective of its study is to draw inference about the population. In other words sampling is only a tool which helps us to know the characteristic of the universe or population by examining only a small part of it.

Simple Sampling: Sampling of attributes may be regarded as the drawing of samples from a population whose members possess the attribute A or not (–A). The choosing of an individual in a sampling is called an event or trial. Possession of a specified attribute A by the individual is called a success.

Simple sampling means random sampling in which each event has the same probability ‘‘p’’ of success and the probability of success is independent of the success or failure of the events in the preceding trials. Thus simple sampling is a special case of random sampling in which trials are independent and probability of success is constant.

Let us consider an example. In an urn there are 2 white and 3 black balls. Probability of drawing a white ball is $\frac{2}{5}$. If a white ball is drawn on the first trial, probability of drawing a white ball will be $\frac{2}{5}$. In the second trial if the ball is not replaced the probability of drawing a white ball will be $\frac{1}{4}$. Hence the sampling though random is not simple.

A random sampling from an infinite population is always a simple since drawing of one individual cannot affect the probability. But a sampling from finite population may or may not be simple.

Let us take a simple sample of ‘n’ members, which will be identical with a series of n independent trials with probability of success p (constant).

The probabilities of 0, 1, 2, …, …, …, n success are the terms of binomial expansion of $(p+q)^n$ which can be written in sequence. The distribution so obtained is called sampling distribution of number successes in the sample. The mean value and S.d (standard error S.E.) are (Missing) np and $\sqrt{npq}$ respectively.

Now proportion of successes in a sample is the number successes divided by the number of members in the sample. Hence,

Mean of proportion of success = p

(S.E) of proportion of success = $\sqrt{\frac{pq}{n}}$

Already we have mentioned that care should be taken in selecting samples, otherwise inference made from the result get will mislead.

There are two types of sampling:

i) Random sampling

ii) Nonrandom sampling.

In random sampling all the items in the population have equal chance of being chosen in the sampling. Whereas in nonrandom sampling personal knowledge and opinion are used to identify the items from the population that are to be included in the Sample. A Sample selected by this sampling is based on someone's expertise. The rigorous statistical analysis that can be done by random sample but cannot be done by non-random sample. Look to the following table for clear understanding the distinction between populations and samples.

	Population	Sample
Definition	Complete enumeration of items is considered	Part of the population chosen for study
Characteristics	Parameters	Statistics
Symbols	Population size = N Population mean = μ Population S.d = σ	Sample size = n Sample mean = $\bar{x}$ Sample S.d = s

In random sample, the probability of an element of the population will or will not be included in the Sample are known. As a result we can

assess mathematically the estimates of the population characteristics from the result of the Sample.

15.3.1 Random and Non-Random Sampling Methods

Sampling is a process of studying the characteristics of a population by selecting samples from it. Let us discuss some random and non-random methods–

Methods of Random Sampling: When a sample is drawn from a population, it is necessary to decide which method should be applicable. The various methods of sampling can be grouped under the heads as **random sampling** and **non-random sampling.**

Some of the sampling methods are as follows:

A) Random Sampling Method:
   i) Simple Random Sampling
   ii) Systematic Sampling
   iii) Stratified Sampling
   iv) Cluster Sampling

B) Non-random Sampling Methods:
   i) Judgment Method
   ii) Quota Sampling
   iii) Convenience Sampling

A) Random Sampling Method:

i) Simple Random Sampling: This method selects samples that allows:

a) Each possible sample to have an equal probability of being selected.

b) Each item in the population has equal probability of being included in the Sample.

Let n units of the sample be drawn from the population one by one, and after drawing the item selected is returned to the population before the next drawing is made. In this case at each stage of drawing, each of N items of the population gets the same probability $\frac{1}{N}$ of being selected.

This sampling procedure is called Simple random Sampling with replacement. In this case, the same unit of the population may occur more than once in the Sample. There will be n possible Samples and each has the probability $\frac{1}{N^n}$.

In other case, the n members of the Sample are drawn one by one, but at any stage members drawn in the previous drawing is not returned to the population. This procedure is known as Simple random Sampling without replacement.

In this case no member of the population can occur more than once in the sample. There are Ncn possible Samples and each Sample has the probability.

$$\frac{n}{N}.\frac{n-1}{N-1}.\frac{n-2}{N-2}.\frac{n-3}{N-3}.---.\frac{1}{N-n+1}=\frac{1}{N_{C_n}}$$

In this case also the probability that any particular population units will be selected of the drawing is:

$$\frac{n-1}{N}.\frac{n-2}{N-1}.\frac{n-3}{N-2}.---.\frac{N-r+1}{N-r+1}.\frac{1}{N-r+1}=\frac{1}{N}$$

Technique of Simple Random Sampling: The easiest way to select a Sample randomly is to use random numbers. These numbers can be taken from a table of random digits or can be generated by a computer programming.

For example, suppose a company has 1000 employees and wish to interview a randomly chosen sample of 10. Write the name of each one on a slip of paper and put the papers on a cylindrical drum and rotate it so that they are thoroughly mixed. Now draw 10 slips at random.

ii) **Systematic Sampling:** In these sampling elements are selected from the population at a uniform interval that is measured in time, space, order or any others.Here each element in the population has the same chance of being selected from the sample. For example suppose we want to interview 20th students on a college campus. We choose a

random starting point in the first 20 names in the student register book and then pick every 20th name thereafter. In this sampling also each element has an equal chance of being selected; but each sample does not have an equal chance of being selected. In this point this sampling differs from Simple random Sampling. In this sampling there is problem of introducing error into the Sample process. The only advantage of this sampling is that it takes less time and lower cost.

iii) **Stratified Sampling:** In this Sampling, population is divided into homogeneous groups. These groups are called strata. There are two approaches:

a) From each strata, select specified number of elements at random.

b) Draw equal number of elements from each strata and give weight to the result according to the Stratum's proportion in the population. In Stratified sampling, every element in the population has chance being selected. In this sampling each group has small variation within itself. There are wide variations between the groups.

iv) **Cluster Sampling:** In this sampling also population is divided into groups and the groups are called Clusters. The assumption is that individual clusters are representative of population. Then random Samples are selected from the clusters. In case of cluster sample there is considerable variation within each group, but the groups are essentially similar to each other. In cluster, stratified and systematic samples attempts are made to approximate simple random sample. These three methods are developed for economical and physical considerations. Simple random sampling is basis of all the samplings and it is fundamental one. Principles of simple random sampling are the foundations of statistical inferences.

B) Non-Random Sampling Methods:

i) Judgment Method: In this method the choice of sample items depends exclusively on the judgment of the investigator. For example, if a sample of 10 artist is to be selected from a group of 50 artists, then the investigator would select 10 artists who, in his opinion, represent the group.

Though this method is simple, it is not scientific because the result may be considerably affected by the personal biasness of the investigator. This method is generally applicable when the size of the sample is small. In general, for solving business problems, making public policy decisions, pilot studies etc we use this method.

ii) Quota Sampling: It is a type of judgment sampling where quotas are set up according to given criteria but within the quotas the selection of sample items depends on personal judgment. For example, if you want to take interview of 100 people, out of which there should be 50 Hindu, 50 Muslim, 50 Sikhs and 50 Christian. Within these quotas, the interviewer is free to select the people interviewed.

iii) Convenience Sampling: In this method the sample is obtained from readily available lists, such as telephone directories, or automobile registration are example of convenience sampling.

Merits of Sampling Method:

1) It is less time consuming compared to census.
2) It is less costly either in terms of money or in terms of man powers, because only a small portion of the total population is studied.
3) More detailed information can be gathered as the size of the sample is small compared to total population.
4) A sample survey generally gives data of better quality than a complete census, because in a sample survey it

may be possible to employ better trained personnel, effect better supervision or use better equipment than is possible or feasible in a complete census.

5) For infinite population, sampling is the only solution.

Limitations of Sampling:

1) Without proper planning, the samples result may be inaccurate and misleading.
2) In absence of qualified persons, the information obtained from samples surveys can not be relied upon.
3) If the information is required from whole population then sampling is meaningless.



CHECK YOUR PROGRESS

Q.1: Fill in the blanks with appropriate words:

i) In ..... sampling all the items in the population have equal chance of being chosen in the sampling.

ii) In ..... sampling personal knowledge and opinion are used to identify the items .

iii) Convenience sampling is a method of ..... sampling.

iv) In ..... Sampling population is divided into homogeneous groups.

v) Sampling is less time consuming compared to .....

15.3.2 Sampling with and without Replacement

Sampling where each member of the population may be chosen more than once is called a sample with replacement, while sampling where each member cannot be chosen more than once is called as sample without replacement.

If we draw an object from an urn, we have the choice of replacing or not replacing the object into the urn before we draw

again. In the first case a particular object can come up again and again, whereas in the second case it can come up only once.

Random Sampling: If in a sampling each member of the population has the same chance of being selected then it is called a random sampling.

A random sample of size 'n' taken with replacement from a population with distribution p(x) = P (X = x) is a collection of random variables x_1, x_2, …, …, ..., x_n each having the same distribution as x and x_1, x_2, …, ..., ..., x_n are independent.

Then μ = Mean of the population = $\frac{\sum x_i p_i}{n}$ = E(n)

Let us denote N numbers of the elements of the set be x_1, x_2, ..., ..., ..., x_n, taking a sample consisting of n elements from the set of N numbers. Sample vatiables are x_1, x_2, ..., ..., ..., x_n.

Mean of population $\mu = \frac{\sum_{i=1}^{N} x_i}{N}$

Mean of sample $\bar{x} = \frac{\sum x_i}{n}$

Now $E(\bar{x}) = E\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(x_i)$ ... ... (1)

$$E(x_i) = \frac{X_1}{N} + \frac{X_2}{N} + ... + \frac{X_n}{N}$$

[As x_i can take any element from X_1, X_2, ..., ..., ..., X_N and its probability is $\frac{1}{N}$]

$\therefore E(x_i) = \frac{1}{N}\sum_{i=1}^{N} x_i = \mu$ ... ... (2)

From (1) and (2): $E(\bar{x}) = \frac{1}{N}\sum_{i=1}^{n} \mu = \frac{1}{N}.n\mu = \mu$

Variance $\sigma = \frac{1}{N}\sum_{i=1}^{N} (X_i - \mu)^2$

If all probabilities are equal and equal to $\frac{1}{N}$ then,

Mean = $\frac{\sum x_i}{n}$ = E(x)

Variance = Var(x) = σ^2 = $E(x^2) - [E(x^2)]^2$

= σ^2

Where $[E(x)]^2 = \sum_{i=1}^{n} x_i^2 p(x_i)$

EXERCISE 15.1

A population consists of five numbers 2, 3, 6, 8 and 11. Consider all possible samples of size 2 that can be drawn with replacement from this population. Find mean and s.d. of the population. Also find mean and s.d. of the samples.

Solution: Let μ = Mean of population

$$= \frac{2+3+6+8+11}{5}$$

$$= \frac{30}{5}$$

$$= 6$$

Again $\sigma^2 = \frac{(2-6)^2+(3-6)^2+(6-6)^2+(8-6)^2+(11-6)^2}{5}$

$$= \frac{16+9+0+4+25}{5} = 10.8$$

Therefore Standard Deviation of the population = σ

$$= \sqrt{10.8}$$

$$= 3.29$$

The samples of size of that can be drawn with replace mean are:

(2, 2)	(2, 3)	(2, 6)	(2, 8)	(2, 11)
(3, 2)	(3, 3)	(3, 6)	(3, 8)	(3, 11)
(6, 2)	(6, 3)	(6, 6)	(6, 8)	(6, 11)
(8, 2)	(8, 3)	(8, 6)	(8, 8)	(8, 11)
(11, 2)	(11, 3)	(11, 6)	(11, 8)	(11, 11)

They are 25 in numbers the corresponding sample means are

2.0	2.5	4.0	5.0	6.5
2.5	3	4.5	5.5	7.0
4.0	4.5	6.0	7.0	8.5
5.0	5.5	7.0	8.0	9.5
6.5	7.0	8.5	9.5	11.0

Mean of the sampling distribution is:

$$\mu_{\bar{x}} = \frac{\text{Sum of the sample means}}{25}$$

$$= \frac{150}{25} = 6.0$$

So $\mu_{\bar{x}} = \mu$

For S.D of samples, subtract 6 from all the means of the sample and square them.

Add all squared members and divide the result by 25.

$$\sigma^2_{\bar{x}} = \frac{135}{25} = 5.4$$

S.D. of samples = $\sigma_{\bar{x}}$

$$= \sqrt{5.4}$$

$$= 3\sqrt{.6} = 2.32$$

Let us consider the same problem without replacement.

Here also, $\mu = 6$

And $\sigma = 3.29$

The samples in this case will be (2,3),(2,6), (2,8), (2,11), (3,6), (3,8),(3,11), (6,8), (6,11) and (8,11).

They are 10 in numbers.

The corresponding sample means are 2.5, 4, 5, 6.5, 4.5, 5.5, 7, 7, 8.5 and 9.5.

Mean of the sampling means are:

$$\mu_{\bar{x}} = \frac{2.4 + 4 + 5 + 6.5 + 4.5 + 5.5 + 7 + 7 + 8.5 + 9.5}{10}$$

$$= \frac{60}{10} = 6$$

So, $\mu_{\bar{x}} = \mu$

Variance of the sample mean is:

$$\sigma^2_{\bar{x}} = \frac{(2.5-6)^2+(4-6)^2+(5-6)^2+(6.5-6)^2+(4.5-6)^2+(5.5-6)^2+(7-6)^2+(8.5-6)^2+(9.5-6)^2}{10}$$

$$= \frac{40.5}{10} = 4.05$$

$$\sigma_{\bar{x}} = 2.01$$

15.3.3 Large Samples

Suppose that a large number n of independents Bernoull's trials is performed and x successes are obtained. The hypothesis we want to test is that the probability of success in each trial is p. Assuming the hypothesis to be correct, the mean and variance of the sampling distribution of the number of success are np and npq.

Again we know for large n, $Z = \frac{x - np}{\sqrt{npq}}$ is distributed as a standard normal variate.

From table $P(|Z| > 3) = 0.0027$. Therefore we conclude that the hypothesis is improbable if the difference between the observed number of successes x and the expected number of successes np, is highly significant.

If $|Z| < 1.96$ the data is consistent with the hypothesis.

If $|Z| > 1.96$ the difference is significant.

Let the observed proportion of success in a sample of n observation be p then,

Mean = p and variance = $\frac{pq}{n}$

Thus $\frac{\frac{x}{n-p}}{\sqrt{\frac{pq}{n}}} \sim N(0, 1)$

This test is valid for large n only.

Standard error: The s.d. of the Sampling distribution of statistic is known as Standard error (SE). If n be the sample size and σ^2 is the population variance then SE of the statistic $\bar{x}$ of the sample is $\frac{\sigma}{\sqrt{n}}$.

SE plays a very important role in the large sample theory and forms the basis of the testing of hypothesis.

If t be any statistic then for large samples,

$$Z = \frac{t - E(t)}{S.E(t)} \sim N(0, 1)$$

Thus if the discrepancy between the observed and expected value of the statistic Z is greater than Z. S.E, then hypothesis is rejected at certain level of significance.

15.4 ERRORS IN SAMPLING PROCEDURE

There are two types of errors in sampling procedure. These are as follows:

1) Sampling error and
2) Non-sampling error

1) Sampling error: A sample covers only a portion of the population from which it is drawn. The information gathered through sample is used for drawing certain inferences about the characteristic of the population. A discrepancy may arise between the population parameter and the sample statistic. The error thus occurs is called sampling error. There are two factors which will lead to sampling error. They are sampling bias and chance factor. When the method of selection is defective, sampling bias will be caused. This error can be minimized by use of proper sampling method. Even when sample bias is absent, discrepancy may occur due to chance.

2) Non-sampling error:

- **Response error:** It is due to response bias- viz vague answer, wrong answer and inaccurate answer given by a respondent.
- **Observational error:** It is due to bias observation.

- **Errors due to non response:** An interviewer may not be able to get information from a respondent.

Test of Significance: A very important aspect of Sampling theory is the study of the test of significance, which enables us on the basis of the Sample results, if the deviation between the observed sample statistic and the hypothetical parameter value.

The technique of randomization used for the selection of Sample units makes the test of significance valid. For applying the test of significance we first set up the hypothesis– a definite statement about the population parameter. This hypothesis is called null hypothesis.



EXERCISE 15.2

A dice is thrown 9,000 times and a throw of 3 or 4 observed 3,240 times. Show that the dice is biased.

Solution: Here, N = 9,000

X = number of successes = 3,240

Hypothesis: Dice is unbiased

P = probability of getting 3 or 4 = $\frac{1}{6}+\frac{1}{6}$

$= \frac{1}{3}$

Expected mean = np = $9{,}000 \times \frac{1}{3}$

= 3,000

n is large $\Rightarrow |Z| = \frac{x-np}{\sqrt{npq}}$

Standard Error (S.E.) = $\sqrt{npq}$

$= \sqrt{9000 \times \frac{1}{3} \times \frac{2}{3}}$

$= \sqrt{2000}$

$= 20\sqrt{5}$

$$Z = \frac{3240 - 3000}{20\sqrt{5}}$$

$$= \frac{240}{20\sqrt{5}}$$

$$= \frac{12\sqrt{5}}{5}$$

$$= 24\sqrt{5}$$

$$= 5.36$$

Since $|Z| > 3$, the dice is biased.

EXERCISE 15.3



A random sample of 500 pineapples was taken from a large contingent and 65 were found to be bad. Show that Standard Error of the population of bad ones in a sample of this size is 0.015.

Solution: Here, n = 500

x = number of bad pineapples in the sample = 65

$$p = \frac{65}{500} = 0.13$$

$$q = 1 - 0.13$$

$$= 0.87$$

$$\text{S.E.} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0.13 \times 0.87}{500}} = 0.015$$



CHECK YOUR PROGRESS

Q.2: In a sample of 1000 people in Guwahati, 840 are rice eaters and rest is wheat eaters. Calculate Z.

.....



15.5 LET US SUM UP

In this unit we have developed the concept of random sampling and discuss simple sample as a particular case of random sample. We have also discussed sampling with and without replacement. In all these discussion we have taken sample from the normal population.

We have also discussed about different methods of sampling - simple random sampling, systematic sampling, stratified sampling and cluster sampling (all these are random sampling). Again we have discussed about some non random sampling like- judgment method, quota method and convenience sampling. We have also discussed about sampling and non sampling errors.



15.6 FURTHER READING

1) Agarwal, D. R. (2006). *Business Statistics*. Delhi: Vrinda Publications.
2) Gupta, S. C. (1994). *Fundamentals of Statistics*. New Delhi: Himalayan Publishing House.
3) Rajagopalan, S. P. & Sattanthan, R. (2009). *Business Statistics and Operations Research*. New Delhi: Tata McGraw-Hill.
4) Sharma, J. K. (2007). *Business Statistics*. New Delhi: Pearson Education Ltd.
5) Verma, A. P. (2007). *Business Statistics*. Guwahat: Asian Books Private Limited.



15.7 ANSWERS TO CHECK YOUR PROGRESS

Ans. to Q. No. 1: i) random sampling, ii) non random sampling, iii) non-random sampling, iv) Stratified sampling, v) census.

Ans. to Q. No. 2: 24.64.



15.8 MODEL QUESTIONS

Q.1: A population is made up of groups that have wide variation within each group but little variation from group to group. What will be the appropriate type of sampling for this population?

Q.2: List the advantages of sampling over census.

Q.3: A population consists of four members 3, 7, 11 and 15. Find all possible samples of size 2 that can be drawn with replacement from this population. Find population mean and population s.d. Find also mean of sampling mean and s.d. of sampling s.d.

Q.4: Solve the above problem without replacement.

Q.5: The masses of 1500 ball bearings are normally distributed with mean 22.40 kg. and s.d. of 0.048 kg. If 3000 random samples of size 36 are drawn from this population, determine the mean and s.d. of the sampling means if sampling is done (a) with replacement (b0 without replacement.

Q.6: In a random sample of 500 people from a particular district, 300 are found to be smokers. In one of 1000 people from another district, 550 are smokers. Calculate Z.

Q.7: A coin is tossed 10,000 times and it turns up head 5,195 times. Can the coin be regarded as unbiased?

Q.8: In some dice throwing experiments, 49,152 times of 75,145 yielded 4, 5, or 6. Is the dice unbiased?

Q.9: A random sample of 500 apples was taken from a large consignment and 60 were found to be bad. Calculate Standard Error.

Q.10: A sample of 900 members has mean 3.4 cm and s.d. 2.61 cm Is the sample from a large population of mean 3.25 cm and s.d. 2.61 cm?

*** ***** ***



REFERENCES

1) Agarwal, D. R. (2006). *Business Statistics*. Delhi: Vrinda Publications.
2) Arora, P. N., Arora, Sumeet & Arora, S. *Comprehensive Statistical Methods.* S. Chand and Company.
3) Beri, G. C. (2010). *Business Statistics* (3rd Ed.). New Delhi: Tata McGraw Hill Education Private Limited.
4) Black, Ken (2007). *Business Statistics for Contemporary Decision Making*. New Delhi: Wiley India Pvt. Ltd.
5) Das, K. K, Bhattacharjee, D. (2008). *A Treatere on Statistical Inference and Distributions*. Asian Books Pvt. Ltd.
6) Feller, W. (1968). *An Introduction to Probability Theory and its Application*. Vol. I. Wiley Eastern.
7) Garrett, H. E. (2005). *Statistics in Psychology & Education*. New Delhi: Paragon International Publishers.
8) Ghosh, R. K. and Saha, S. *Business Mathematics and Statistice.* New Central Book Agency (P) Ltd.
9) Goon, A.M., Gupta, M. K. and Dasgupts, B. (1997). *An Ouline of Statistical Theory*. Vol. I. World Press.
10) Goon, A. M., Gupta. M. P. and Dasgupta, B. (2002). *Fundementals of Statistics* (Vol. I). Seventh Revised Edition. Kolkata: The World Press Pvt. Ltd.
11) Gupta, S. C. (1994). *Fundamentals of Statistics*. New Delhi: Himalayan Publishing House.
12) Gupta, S. C. and Gupta, Indira. *Business Statistics.* New Delhi: Himalaya Publishing House.
13) Gupta, S. C., Kapoor, V. K. (2009). *Fundamentals of Mathematical Statistics*. Sultan Chand & Sons.
14) Gupta, S. K. (1968). *Elements of Probability*. NCERT.
15) Kakoty, S. C. (2003). *Mathematical Statisties: Theory and Applications*. Dibrugarh, Assam: Kasturba Prakashass.

16) Medhi, J. (2005). *Statistical Method*. New Age International Publishers.

17) Mukherjee, Kalyan Kumar (1993). *Probability and Statistics*. New Central Book Agency.

18) Mukhopadhyay, P. (1991). *Theory of Probability*. New Central Book Agency (P) Ltd.

19) Mukhopadhyay, P. (1996). *Mathematical Statistics*. New Central Book Agency (P) Ltd.

20) Murray, R., Spiegel, Jhon, Schiller, R., Srinivasan, Alu (2005). *Probability and Statistics*. Schaum's Outline Series.

21) Rajagopalan, S. P. & Sattanthan, R. (2009). *Business Statistics and Operations Research*. New Delhi: Tata McGraw-Hill.

22) Rao. G. Shanker, *Probability and Statistics*. Universitis Press.

23) R. S. N. Pillai and Bagavathi. *Statistics*. New Delhi: S. Chand and Company Ltd.

24) Saxena, H. C. (1994). *Probability and its Application*. Sixth Edition. New Delhi: S. Chand and Company.

25) Sharma, R. D. (2009). *Mathematics for Class X*. New Delhi: Dhanpat Rai Publications.

26) Sharma, J. K. (2007). *Business Statistics*. New Delhi: Pearson Education Ltd.

27) Shukla, S. M. & Sahai, S. P. (2008). *Business Statistics Theory & Practice*. Agra: Sahitya Bhawan Publications.

28) Singh, R. (2012). *An Introduction to Probability and Probability Distributions*. Books and Ellied (P) Ltd.

29) Srivastava, S. C. & Srivastava, S. (2009). *Fundamentals of Statistics*. New Delhi: Anmol Publications Pvt. Ltd.

30) Verma, A. P. (2007). *Business Statistics*. Guwahat: Asian Books Private Limited.

Name of the Paper: .....
Course Code: .....

Centre for Internal Quality Assurance (CIQA)
Krishna Kanta Handiqui State Open University
City Office: Housefed Complex, Guwahati-781006

Learner's Feedback on Course

Dear Learner,

Regarding the course as mentioned above, we would like to know your opinions and comments so as to improve the quality of self learning materials in future. Please respond to the following statements by ticking the number you feel most reflect your opinion. After completion of the additional comments, please detach the page and send/mail the same to us at the address given below.

The Director, Centre for Internal Quality Assurance, KKHandiqui State Open University
Housefed Complex, Dispur, Guwahati-781006
(E-mail id: ciqa@kkhsou.in)

1) Approximately how many hours did you spend for studying the units in the course? ☐☐☐

2) Please give your opinions (by ✓ mark) to the following items based on your reading of the block:

Sl. No.	Statements	Strongly Agree	Agree	Neither Agree nor Disagree	Disagree	Strongly Disagree
I)	The SLMs of the course fulfil the learning objectives set out in the beginning of each unit					
II)	The units of the SLM could satisfy my academic needs and expectations					
III)	The Unit writers have excellent knowledge about the course contents					
IV)	Language and contents of the units were presented at a level which I could readily understand					
V)	Ample opportunity for participation in the activities provided in the units					
VI)	Used enough Illustrations (Diagrams, tables etc.) for conceptual clarity					
VII)	Quality of content is engaging, relevant, and up-to-date					
VIII)	The self check questions are very helpful					
IX)	The Possible/model questions and the answers to check my progress have benefited me a lot					

Additional Comments: *(Please feel free to provide your open comments)*

1) Which aspects of the SLM, according to you, worked well?

   .....

   .....

2) What sort of changes/improvements do you feel KKHSOU could implement to improve the overall quality of the SLM?

   .....

   .....

Thank you for taking the time to complete this form.