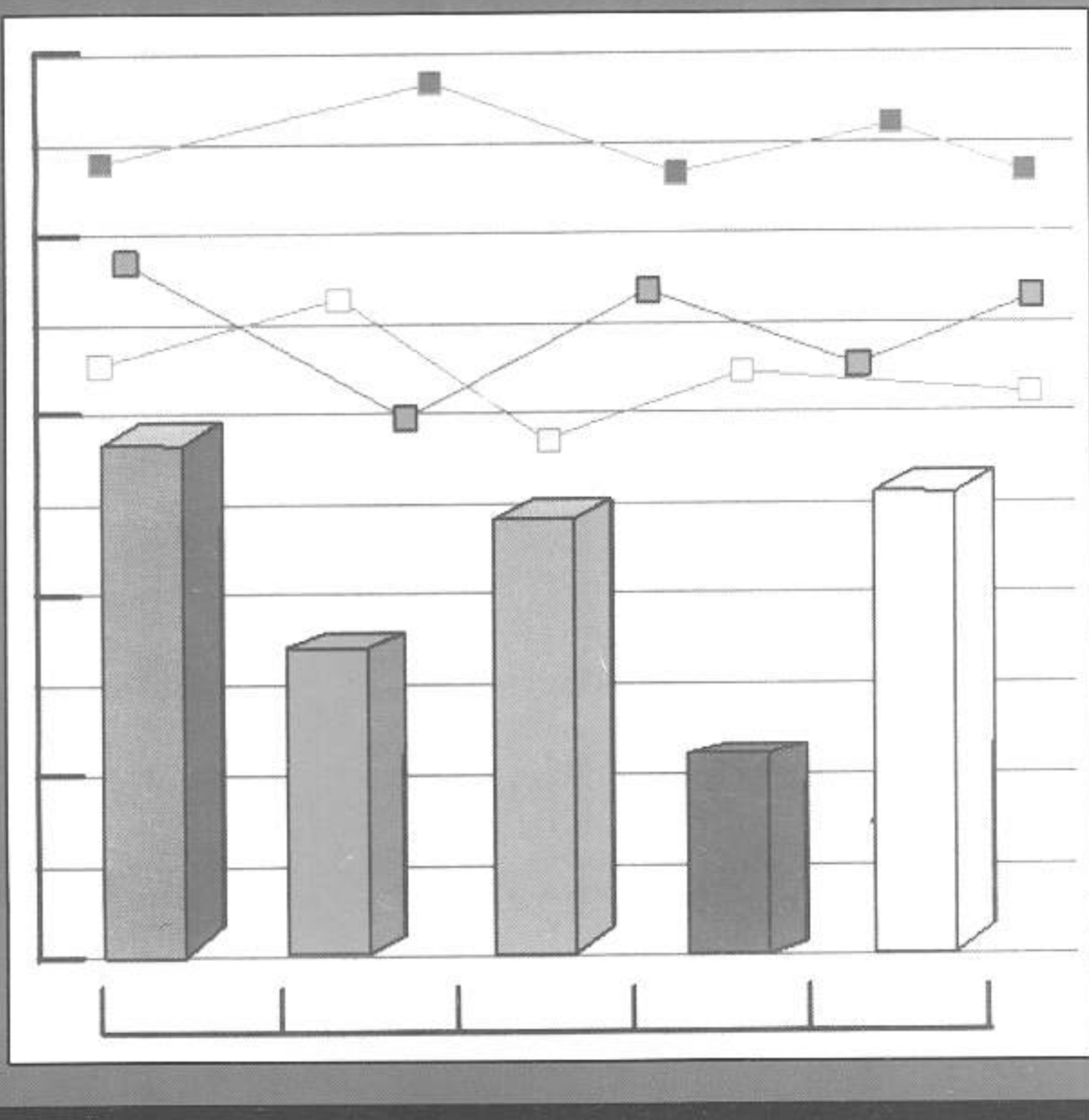
GAUHATI UNIVERSITY

Institute of Distance and Open Learning

ECO-03-4

Statistical Methods for Economic Analysis



www.idolgu.in

E-Portal:www.bodhidroom.idolgu.in

ECO-03-4

Institute of Distance and Open Learning
Gauhati University

M.A./M.Sc. in Economics
Semester 1

Paper IV
Statistical Methods for Economic Analysis

Contents:

Course Coordinator :

Prof. Pranab Jyoti Das : Director, i/c GU IDOL

Dipankar Saikia : Editor, SLM, GU IDOL

Contributors :

Miss Arati Bharali : Research Scholar
Department of Economics
Gauhati University

Editorial Team :

Dr. Ratul Mahanta : Co-ordinator, IDOL, Gauhati University
Department of Economics, Gauhati University

Daisy Konwar : Asstt. Professor in Economics,
IDOL, Gauhati University

Dipankar Saikia : Editor, SLM, IDOL, Gauhati University

Cover Page Designing :

Bhaskarjyoti Goswami : IDOL, Gauhati University

Reviewer :

Rantu Hazarika : Research Scholar
Dept. of Economics
Dibrugarh University

Reprint : July, 2017

MA/M.Sc. Economics
Institute of Distance and Open Learning
GAUHATI UNIVERSITY

COURSE STRUCTURE

A student shall do a total number of sixteen papers in the four Semesters. Each paper will carry 100 marks - 20 marks for internal evaluation during the semester and 80 marks for external evaluation through end semester examination. All the papers in the First, Second and Third Semesters will be compulsory. The paper XIII and XIV of the Fourth Semester will also be compulsory. The remaining two papers for the Fourth Semesters will be chosen by a student from the optional papers. The names and numbers assigned to the papers are as follows.

First Semester

- I Microeconomic Theory
- II Macroeconomic Theory - I
- III Mathematical Methods for Economic Analysis-I
- IV Statistical Methods for Economic Analysis

Second Semester

- V Advanced Microeconomics
- VI Macroeconomic Theory -II
- VII Mathematical Methods for Economic Analysis-II
- VIII Elementary Econometrics

Third Semester

- IX Development Economics-I
- X International Economics
- XI Issues in Indian Economy
- XII Public Finance-I

Fourth Semester

- XIII Development Economics-II
- XIV Public Finance-II

Papers XV and XVI are optional

A student has to choose from any the following courses.

- XV (a) Population and Human Resource Development or
- XV (b) Econometric Methods
- XVI (a) Environmental Economics or
- XVI (b) Financial System

Paper Introduction

Economics is defined as the study of how people behave with regard to the production and consumption of goods and statistics is the subject that helps in collection, analysis and interpretation of data. It helps in simpler data presentation in an effective manner so as to conclude the result even to a person that have no knowledge about that topic. Thus, statistics helps in giving an easy to understand quantitative expression to economic problems, for eg., if one says that there are many unemployed people in India, it does not give you any idea, whereas if one says that about 25% people of India are unemployed, you get a better idea.

Statistical methods are used to prepare national income accounts, index numbers, to forecast about future economic consequences of policies undertaken today, etc. Therefore a sound knowledge of statistical tools and techniques and their application in economics is very essential for the learners. This paper contains 4 units.

The first unit deals with the concept of probability. After reading the unit you will be able to get an overall idea about what actually probability is, the various related concepts like basic probability rules, conditional probability, Random variable, Mathematical expectations etc. You will also learn about moments relating to discrete and continuous random variables.

The second unit can be actually said as an extension of the first unit. This unit is about 3 different probability distributions functions viz., binomial distribution, poisson distribution and normal distribution. After going through this unit you will be able to differentiate between discrete and continuous probability distribution function. You will also learn about their application. The unit also tries to give a basic idea about Moment generating function and the central limit theory.

In the third unit, you will learn about various concepts regarding Income distribution. The unit tries to introduce to you the concepts of Pareto' law of Income distribution, log-normal distribution, different measures of income distribution etc. It gives a special emphasize on Gini coefficient and Lorenz curve.

The last unit is about index numbers. After going through this unit you will get an idea about the index numbers of Laspayre, Paasche, Fisher etc. and the inter-relation between them. You will also know about different tests used to judge the adequacy of index numbers viz., the time reversal, factor reversal and circular test etc. You will be able to know about base shifting, splicing and deflating index numbers. Finally, the unit also introduces the concept of indices of industrial production. You will get to know what these indices are.

●●●

Unit-1 : Probability Theory

Contents :

1.1 Introduction :

Probability is the measure of likelihood of occurances of chance events. Events whose occurances entirely depend upon factors beyond human control are called chance events. Thus, probability gives a quantitative measure of the chance of a random experiment.

The theory of probability has its origin in the games of chance related to gambling. A systematic and scientific foundation of the mathematical theory of probability was laid in mid-seventeenth century by the French mathematicians Blaise Pascal (1623-62) and Pierre de Fermat (1601-65). Thomas Bayes (1702-61) introduced the concept of inverse probability. Russian mathematicians have made great contributions to the theory of probability. The modern theory of probability was developed by Russian mathematicians like Chebychev (1821-94), A Markov (1856-1922) and A.N. Kolmogorov.

There are three approaches to probability. They are :

(i) Classical Approach

(ii) Emperical Approach

(iii) Axiomatic Approach

1.2 Objectives :

This unit is designed to help you understand the concept of probability and its related ideas. After reading this unit you will be able to.

(a) Explain different concepts related to probability.

(b) Analyse the pattern of distribution on the basis of moments.

(c) Draw conclusion of any event on the basis of mathematical expectation.

(d) Formulate any probabilistic decision in life.

1.3 Definition to the Theory of Probability:

The axiomatic approach to probability, which closely relates the theory of probability with the modern metric theory of functions and also set theory, was propounded by a Russian mathematician A.N. Kolmogorov in 1933. The axiomatic of probability includes both the classical and statistical definitions as particular cases and overcomes the deficiencies of each of them. On the basis of 3 fundamental axioms and relying entirely on the basis of deduction Kolmogorov was able to bring together the diverse strands of probability into a unified axiomised system. The axioms provide a set of rules which define relationship between abstract entities More precisely, under axiomatic approach, the probability can be deduced from mathematical concepts.

Prior to 1933, there were three main approaches to the theory of probability. Historically, the oldest and simplest way of measuring probabilities is the classical approach. It applies when all possible outcomes of an experiment are equally likely and mutually exclusive. Thus, if an experiment results in 'N' exhaustive, mutually exclusive and equally likely outcomes and 'm' of them are favourable to the occurrence of an event 'A' then the probability P (also called its success) of the occurance of the event A is given by,

$$p = P(A) = \frac{\text{Favourable number of cases}}{\text{Exhaustive number of cases}}$$

$$= \frac{m}{N}$$

The probability of non occurance of the event (also called its failure) is given by,

$$q = P(\bar{A}) = \frac{N - m}{N}$$

$$= 1 - \frac{m}{N}$$

$$= 1 - p$$

$$\Rightarrow p + q = 1$$

Here, $0 \leq p \leq 1$

and $0 \leq q \leq 1$

As m and N are non negative integers, $P(A) \geq 0$. Again, as the favourable number of cases for A are always less than or equal to the total number of cases N i.e. $m \leq N$; hence $P(A) \leq 1$. For any event A, if $P(A) = 0$, then A is impossible or null event. Again, if $P(A) = 1$, then A is a certain event.

The classical approach is stemming from its reliance on outcomes which are equally likely as well as finite in number. To overcome this limitation, the frequency interpretation of probability has been evolved according to which the probability of an event (outcome or happening) is the proportion of time that events of the same kind will occur in the long run. According to Von Misses, "If a trial is repeated a number of times under essentially homogenous and identical conditions, then the limiting value of the ratio of the number of times the event happens to the number of trials, as the number of trials becomes indefinitely large, is called the probability of happening of an event." Symbolically, if an event A, occurs 'm' times in a series of 'n' Trials, then the probability p of happening of event A is given by,

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

STOP TO CONSIDER

Terminologies related to Probability Theory :

1. Random Experiment : If in each trial of an experiment conducted under identical conditions, the outcome is not unique, but may be any one of the possible outcomes, then such an experiment is called a random experiment. Examples of random experiments are : tossing a coin, throwing a die, selecting a card from a pack of cards etc.

2. Outcome : The result of a random experiment is called an outcome.

3. Trial and event : Any particular performance of a random experiment is called a trial and outcome or combinations of outcomes are termed as events.

4. Exhaustive Events or Cases : The total number of possible outcomes of a random experiment is known as the exhaustive events or cases.

5. Favourable Events or Cases : The number of cases favourable to an event in a trial is the number of outcomes which entail the happening of the event and are known as the number of favourable events.

6. Mutually Exclusive Events : Events are said to be mutually exclusive or incompatible if the happening of any one of them precludes the happening of all the others, i.e., if no two or more of them can happen simultaneously in the same trial.

7. Equally Likely Events : Outcomes of trial are said to be equally likely if taking into consideration all the relevant evidences, there is no reason to expect one in preference to the others. (or in simple terms, when the probability of happening of both the events are same).

8. Independent Events : Several events are said to be independent if the happening (or non-happening) of an event is not affected by the happening (or non-happening) of the remaining events.

The third main approach is subjective approach was propounded by J.M. Keynes and L.T. Savage. This approach interpreted probabilities as measuring the strength of one's beliefs or confidence in the occurance of a particular event.

When an experiment is performed, it generates certain outcomes. The outcomes are called sample points or elementary events. A collection of all possible outcomes or sample points is called sample space. It is generally denoted by the leter S. For example, tossing an unbiased coin have two sample points is head (H) and tail (T). The sample space will be,

$S = \{H, T\}$

If A and B indicates the events of obtaining a head and a tail respectively, they are mutually exclusive events. i.e. outcome of one event restricts the outcome of the other event. It can be denoted with the help of a Venn diagram.

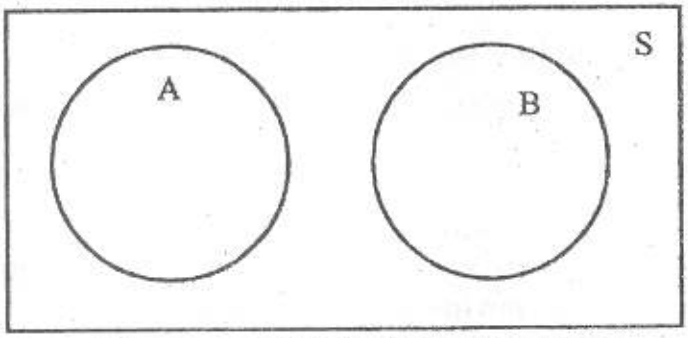


Diagram : Mutually Exclusive Events

This is an example of finite outcomes. In cases of large or infinite number of outcomes, corresponding sample spaces are best described as a statement or rule. Thus, if S is the set of odd positive numbers, we write,

$S = \{2k + 1 \mid k = 0, 1, 2, ...\}$

Thus, according to the axiomatic approach, probability is defined as a function defined on events (Subsets of S). This set function assigns to each event A, a certain real number P(A) which satisfies the following three axioms.

Axiom 1 : (Axiom of Non-negativity)

The probability of an event is a non negative real number i.e.,

$P(A) \geq 0$ for any subset A of S.

Axiom 2 : (Axiom of Certainty)

The probability of a certain event is equal to unity. i.e.

$P(S) = 1$

Axiom 3 : (Axiom of Additivity)

If $A_1, A_2, A_3,...$ is a finite or infinite sequence of mutually exclusive events of S, then

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

i.e. the probability of the union of mutually exclusive events is the sum of the probabilities of the individual events.

As postulated above, the axioms of probability apply only when the sample space S is discrete, i.e. contains a finite number of elements or an infinite though countable number of elements.

Let us understand,

Generally, we talk about probability as a fraction, a decimal or a percent; for instance,

- If we toss a coin the probability of getting a head is 1/2.
- The probability that a baseball player will get a hit is 0.273.
- The probability that it will rain today is 20%.

1.3.1 Basic Probability Rules :

The following important laws of probability follow immediately from the above given axioms :

I : If A is an event in a discrete sample space S, then the probability of occurance of event P(A) is equal to the sum of the probabilities of the individual outcomes comprising A.

Thus, if $0_1, 0_2, 0_3$ ... be the finite or infinite sequence of outcomes which comprise the event A then,

$$P(A) = P(0_1) + P(0_2) + P(0_3) + ...$$

Example ,

If a balanced coin is tossed twice, what is the probability of obtaining a tail on the first coin ?

Solution :

Here, the Sample Space (S) = {HH, HT, TH, TT}

If A denotes the probability of obtaining a tail on the first coin, then the required event is,

$$A = \{TH, TT\}$$

$$P(A) = P(TH) + P(TT)$$

$$= \frac{1}{4} + \frac{1}{4} = \frac{2}{4} = \frac{1}{2} = 0.5$$

II : If an experiment can result in any one of 'N' equally likely and mutually exclusive outcomes, and if 'n' of these outcomes together constitute event A, then,

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

Thus, the classical concept of probability can be derived from the three axioms in the special case when the individual outcomes of an experiment are all equally likely.

Example :

From a bag containing 6 green and 3 brown balls, 2 balls are drawn at random. What is the probability that both the balls drawn are green?

Solution:

2 balls can be drawn out of 9 balls in

$^9C_2 = \frac{9!}{2!\,3!} = 36$ ways which are all equally likely.

2 green balls can be drawn out of 6 green balls in $^6C_2 = \frac{6!}{2!\,4!} =$ 15 ways.

Hence the required probability,

$$\frac{15}{36} = \frac{5}{12}$$

III : If A and A′ are complementary events in a sample space then,

$$P(A') = 1 - P(A) \text{ and } \quad P(A) = 1 - P(A')$$

The complement of event A is denoted by A′. A′ is the event in S which consists all the elements of S that are not included in A.

Example :

If we consider the experiment of tossing an unbiased coin, the sample space is

$S = \{H, T\}$

If A denotes the event of having head (H) in the toss, then,

$A = \{H\}$

Then the complementary event A′ denotes the event of having tail (T), then

$A' = \{T\}$

Hence, $P(A) = 1 - P(A') = 1 - \frac{1}{2} = 0.5$

IV : P(ϕ) = 0 for any sample space S, where ϕ denotes the null or empty set. This shows that the probability of an event which is sure not to occur is zero.

Example :

If we consider the experiment of rolling an unbiased die, the sample space is

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

If A denotes the event of drawing a face numbered 7, then

$A = \{7\}$

Hence, $P(A) = \phi = 0$

V : If A and B are events in a sample space S, and $A \subset B$ then $P(A) \leq P(B)$

Example : If A denotes the event of drawing a heart from an ordinary pack of 52 cards, then

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

If B denotes the event of drawing a red card from the same ordinary pack, then

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

Here, $A \subset B$

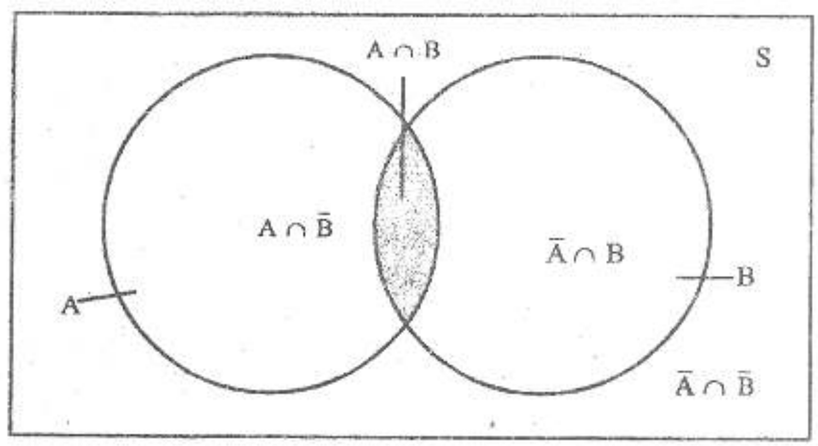
and $P(A) \leq P(B)$

VI : $0 \leq P(A) \leq 1$ for any event A in the sample space S.

VII : For the third postulate to be applicable, all the events $A_1, A_2, A_3, \ldots$ must all be mutually exclusive. Hence it is sometimes referred to as a special addition rule. For events which are mutually exclusive, the more general additive rule applies.

But, If A and B are any two events in a sample space S, which are not mutually exclusive, then

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



We have,

$$A \cup B = A \cup (\bar{A} \cap B)$$

$$P(A \cup B) = P\{A \cup (\bar{A} \cap B)\}$$

$$= P(A) + P(\bar{A} \cap B)$$

[Since A and $(\bar{A} \cap B)$ are mutually exclusive events]

$$= P(A) + P(\bar{A} \cap B) + P(A \cap B) - P(A \cap B)$$

$= P(A) + P\{(\bar{A} \cap B) + (A \cap B)\} - P(A \cap B)$

[$\because$ $(\bar{A} \cap B)$ and $(A \cap B)$ are mutually exclusive events]

$= P(A) + P(B) - P(A \cap B)$ [$\because P\{(\bar{A} \cap B) + (A \cap B)\} = P(B)$]

Remark :

When the event A and B are mutually exclusive events, then

$P(A \cap B) = 0$ and hence

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

The above additive rule is also known as the theorem of total probability. This rule can be generalised. Thus, if the events $A_1, A_2, ...,$ A_n are not mutually exclusive, then

$$P(A_1 + A_2 + ... + A_n) = \sum_{i=1}^{n} P(A_i) - \sum_{i,j=1}^{n} P(A_i A_j) + ... + (-1)^{n-1} P(A_1, A_2 .... A_n)$$

SELF-ASKING QUESTION

Try yourself to prove all the basic probability rules with the help of simple examples from day to day live.

Permutations :

A permutation is an arrangement of objects where order is important. In general, the number of permutations can be derived from multiplication principle.

Formula for computing the number of permutations of r objects choosen from n distinct objects is $r \leq n$. The notation for these permutations is $P(n, r)$ and the formula is :

$$P(n, r) = n.(n-1)(n-2) ..... [n-(r-1)]$$

We often use factorial notation to rewrite this formula,

$$P(n, r) = \frac{n!}{(n-r)!}$$

Combinations :

A combination is an arrangement of objects in which order is not important. Here also $r \leq n$. Notation to represent combination $C(n, r)$ formula is,

$$C(n, r) = \frac{n!}{(n-r)!\,r!}$$

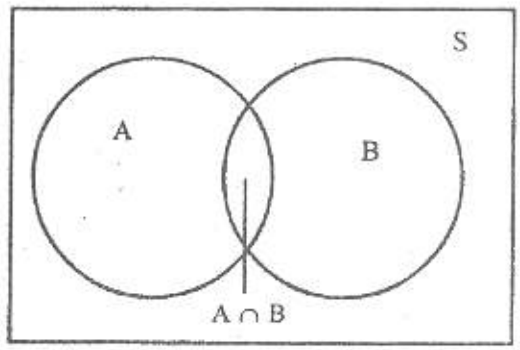
N.B.

Factorial n (= $|\underline{n}$) or n factorial (= n!) can be written as,

$|\underline{n} = n! = 1 \times 2 \times 3 \times .... \times (n-1) \times n$

1.4 Conditional Probability :

Two events A and B are said to be dependent when B can occur only when A is known to have occurred (or vice-versa). The probability attached to such an event is called the conditional probability. We sometime encounter situations where we are to estimate the probability of occurance of an event A knowing that another event B had already accrued. It is denoted by $P(A \mid B)$ and is read as the conditional probability of A given that B has already occurred. $P(B \mid A)$ can be similarly interpreted.



Suppose, we have a total of N outcomes in the sample space S of which ^nA outcomes are favourable to the event A while ^nB are favourable to the event B. Further let ^nAB be the outcomes favourable to both the events A and B i.e. the compound event AB. From the above Venn diagram it is clear that out of ^nB outcomes favourable to the event B, ^nAB outcomes are also favourable to the event A. Hence,

$$P(A \mid B) = \frac{^nAB}{^nB}$$

$$= \frac{^nAB}{N} \div \frac{^nB}{N}$$

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

Similarly, it can be proved that,

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

It can be noted that conditional probabilities $P(A \mid B)$ and $P(B \mid A)$ are defined if and only if $P(B) \neq 0$ and $P(A) \neq 0$.

Example :

If a card drawn from a pack of cards is red, what is the probability that it is the queen of diamond?

Solution :

Let A denotes the event that the card drawn is red and B is the event that it is the queen of diamond.

Hence, $P(A) = \frac{26}{52}$

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

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{1/52}{26/52} = \frac{1}{26}$$

Notice the following,

$$P(A \cap B) = P(A)\, P(B \mid A)$$
$$= P(B)\, P(A \mid B)$$

In other words, the probability of the simultaneous occurance of any two events A and B (in the sample space S) is equal to the probability of B multiplied by the conditional probability of A given that B has already occurred. It is also equal to the probability of A multiplied by the conditional probability of B given that A has occurred.

This rule is known as multiplicative law of probability or the theorem of compound probability. The multiplicative law can be extended to any number of events.

Thus, for n events $A_1, A_2, \ldots, A_n$ we have,

$$P(A_1 \cap A_2 \cap \ldots \cap A_n)$$
$$= P(A_1)\, P(A_2 \mid A_1)\, P(A_3 \mid A_1 \cap A_2) \ldots P(A_n \mid A_1 \cap A_2 \cap \ldots \cap A_{n-1})$$

STOP TO CONSIDER

When we know that a particular event B has occurred, instead of S, we concentrate our attention on B only and the conditional probability of A given B will be analogously the ratio of the probability of that part of A which is included in B (i.e., $A \cap B$) to the probability of B. It therefore, reflects the change of viewpoint only, namely, instead of S we have to concentrate on B only.

1.4.1 Independent Events :

Events are said to be independent of each other if happening of any one of them is not affected by and does not affect the happening of any one of the other.

If A and B are independent events so that the probability of occurance or non-occurance of A is not affected by occurance or non-occurance of B, then we have

$P(A \mid B) = P(A)$ and $P(B \mid A) = P(B)$

Hence, if the two events A and B are independent the multiplicative rule reduces to.

$$P(A \cap B) = P(B)\,P(A \mid B) = P(B)\,P(A)$$

and $P(A \cap B) = P(A)\,P(B \mid A) = P(A)\,P(B)$

This is multiplicative law of probability for independent events. Thus, two events A and B are independent if and only if,

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

Corollary I :

It can be easily deduced that if A and B are independent, then

(a) $\bar{A}$ and B are independent.

(b) A and $\bar{B}$ are independent.

(c) $\bar{A}$ and $\bar{B}$ are independent.

Corollary II :

If $A_1, A_2, ..., A_n$ are 'n' independent events with respective probabilities of occurance, $P(A_1), P(A_2), ..., P(A_n)$ respectively, then the probability that at least one of the 'n' events occur is given by,

$$P(A_1 \cup A_2 \cup ... \cup A_n)$$
$$= 1 - [\{1 - P(A_1)\}\{1 - P(A_2)\}.... \{1-P(A_n)\}]$$
$$= 1 - [P(\bar{A}_1)\,P(\bar{A}_2)\,....P(\bar{A}_n)]$$

1.4.2 Pairwise Independence and Mutual Independence :

Let, $A_1, A_2, ...A_n$ be 'n' events associated with sample space S. They are said to be pairwise independent if,

$$P(A_i \cap A_j) = P(A_i)\,P(A_j);\ \forall\ i \neq j = 1, 2, ...n$$

On the other hand, $A_1, A_2, ... A_n$ are mutually independent if,

(1) $P(A_i \cap A_j) = P(A_i)\,P(A_j)\ ;\ i \neq j = 1, 2, ... n$

(2) $P(A_i \cap A_j \cap A_k) = P(A_i)\,P(A_j)\,P(A_k)\ ;\ i \neq j \neq k = 1, 2, ..n$

(n-1) $P(A_1 \cap A_2 \cap ... \cap A_n) = P(A_1)\,P(A_2)\ ....\ P(A_n)$

Thus, the total number of conditions for the mutual independence of the events $A_1, A_2, ... A_n$ events are,

$$ {}^nC_2 + {}^nC_3 + ... + {}^nC_n = ({}^nC_0 + {}^nC_1 + ... + {}^nC_n) - {}^nC_0 - {}^nC_1 $$
$$ = 2^n - 1 - n $$

In particular for three events A_1, A_2 and A_3 (N=3) we have the following $2^3 - 1 - 3 = 4$ conditions for their mutual independence.

$$P(A_1 \cap A_2) = P(A_1)P(A_2)$$
$$P(A_2 \cap A_3) = P(A_2)P(A_3)$$
$$P(A_1 \cap A_3) = P(A_1)P(A_3)$$
$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3)$$

Hence, events are mutually independent if they are independent by pairs, and by triplets and by quadruples and so on. Again, mutual independence implies pair wise independence though the converse is not necessarily true.

1.5 Baye's Theorem :

Baye's theorem is based on the formula for conditional probability. If $E_1, E_2, ... E_n$ are mutually exclusive events with $P(E_i) \neq 0$, (i = 1, 2, ..., n) then for any arbitrary event A which is a subset of $\bigcup_{i=1}^{n} E_i$ such that $P(A) > 0$,

We have,

$$P(E_i|A) = \frac{P(E_i)P(A|E_i)}{\sum_{i=1}^{n} P(E_i)P(A|E_i)}$$

(1) The probabilities $P(E_1)$, $P(E_2)$, ... $P(E_n)$ are referred to as 'apriori' probabilities, since their values are already known even before the actual experiment is carried out.

(2) The conditional probabilities $P(A | E_i)$, i = 1, 2, ..., n are termed as likelihoods since they give us the possible ways in which the event A can occur, given the occurance of the events $E_1, E_2, ..., E_n$.

(3) The probabilities $P(E_i | A)$, i = 1, 2, ..., n are known as the 'posteriori' probabilities since they can be estimated only after the results of the experiment are known.

Hence, the Baye's Theorem is used to obtain the 'posteriori' probabilities.

$P(E_i|A)$, $(i = 1, 2, ..., n)$, that is the probabilities of the events $E_1, E_2, ... E_n$ knowing that the event A has occurred in the experiment

SELF-ASKING QUESTION

Do you think there is any difference between general conditional probability and Bayes' Theorem? Justify your answer.

1.6 Solved Examples :

Example 1. An urn contains 8 red, 3 white and 9 black balls. If 3 balls are drawn at random, determine the probability of the event that,

(i) All three balls are red,

(ii) All three balls are white,

(iii) Two are red balls and one is black ball.

[G.U. (MA/MSc.) '95]

Solution :

3 balls can be drawn out of 20 balls in $^{20}C_3 = 1140$ ways.

Which are equally likely and mutually exclusive events.

(1) 3 red balls can be drawn out of 8 red balls in $^8C_3 = 56$ ways.

Hence, the required probability that all three balls are red is,

$$\frac{^8C_3}{^{20}C_3} = \frac{56}{1140}$$

(ii) 3 white balls can be drawn out of 3 white balls in $^3C_3 = 1$ way.

Hence, the probability that all three balls are white is,

$$\frac{^3C_3}{^{20}C_3} = \frac{1}{1140}$$

(iii) 2 red balls can be drawn out of 8 red balls in $^8C_2$ ways. One black ball can be drawn out of 9 black balls in $^9C_1$ ways.

Hence, 2 red balls and 1 black ball can be drawn in

$$^8C_2 \times {}^9C_1 \text{ ways.}$$

Hence, the probability that 2 balls are red and 1 is black is given by,

$$\frac{^8C_2 \times {}^9C_1}{^{20}C_3} = \frac{28 \times 9}{1140} = \frac{252}{1140}$$

Example 2. Two digits are selected at random from the digits 1 through 9. If the sum is even, find the probability P that both are odd.

[G.U. (MA/MSc) '96]

Solution :

There are 9 digits i.e. 1, 2, 3, 4, 5, 6, 7, 8 and 9 of which 5 are odd viz. 1, 3, 5, 7, 9 while 4 are even viz 2, 4, 6, 8.

Two digits can be drawn at random out of 9 digits in ${}^9C_2$ ways which are equally likely and mutually exclusive events.

The sum of the digits drawn will be even if either both the digits drawn are even or both of them are odd.

Now, let E denotes the event that sum of the two drawn digits is even.

E_1 be the event that both the digits drawn are even, E_2 be the event that both the drawn digits are odd.

Hence,

$$P(E_1) = \frac{{}^4C_2}{{}^9C_2} = \frac{1}{6}$$

$$P(E_2) = \frac{{}^5C_2}{{}^9C_2} = \frac{5}{18}$$

By the addition theorem of probability

$$P(E) = P(E_1) + P(E_2)$$

$$= \frac{1}{6} + \frac{5}{18} = \frac{8}{18}$$

Hence, the conditional probability that both the digits are odd, given that the sum is even is given by,

$$P(E_2 | E) = \frac{P(E_2 \cap E)}{P(E)}$$

$$= \frac{P(E_2)}{P(E)} \; (\because E_2 \text{ is a subset of } E)$$

$$= \frac{5}{18} \div \frac{8}{18}$$

$$= \frac{5}{8}$$

Example 3. If for a mutually exclusive and exhaustive events A and B, P(B) = 2 P(A) and $A \cup B = S$, Find P(A).

[G.U. (MA/MSc) '97]

Solution : Given, $A \cup B = S$

$$\Rightarrow P(A) \cup P(B) = P(S)$$

$$\Rightarrow P(A) + P(B) = 1 \; [\because A \text{ and } B \text{ are MEE}]$$

$$\Rightarrow P(A) + 2P(A) = 1$$

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

Example 4. If $P(A) = \frac{1}{2}$, $P(B) = \frac{1}{3}$ and $P(AB) = \frac{1}{6}$. Find $P(A + B)$

[G.U. (MA/MSc.) '99]

Solution :

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

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

Example 5. Suppose, A and B are two events and that $P(A) = P_1$, $P(B) = P_2$ and $P(A \cap B) = P_3$

Find (i) $P(A \cup B)$ (ii) $P(\bar{A} \cap \bar{B})$ (iii) $P(A \cap \bar{B})$
(iv) $P(\bar{A} \cap B)$ (v) $P(A \mid B)$

[G.U. (MA/MSc.) '99]

Solution :

The two events are not mutually exclusive.

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

$= P_1 + P_2 - P_3$

(ii) $P(\bar{A} \cap \bar{B}) = P(\overline{A \cup B})$

$= P(S) - P(P_1 + P_2 - P_3)$

$= 1 - P_1 - P_2 + P_3$

(iii) $P(A \cap \bar{B}) = P(A) - P(A \cap B)$

$= P_1 - P_3$

(iv) $P(\bar{A} \cap B) = P(B) - P(A \cap B)$

$= P_2 - P_3$

(v) $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{P_3}{P_2}$

Example 6. In a simple throw of two dice, find the probability of obtaining (i) A sum of 9 on both the dice (ii) A sum less them 8.

[G.U. (MA/MSc) '02]

Solution :

Total number of cases $= 6^2 = 36$

(i) Favourable number of cases = (3, 6) , (6, 3), (5, 4), (4, 5).

$\therefore$ Required probability $= \frac{4}{36} = \frac{1}{9}$

(ii) Favourable number of cases =

{(1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (3, 1), (3, 2), (3, 3), (3, 4), (4, 1), (4, 2), (4, 3), (5, 1), (5, 2), (6, 1)}

$\therefore$ Required Probability $= \frac{20}{36} = \frac{5}{9}$

Example 7. An unbiased coin is tossed 3 times. If A is the event that a head occurs in each of the first two tosses, B is the event that a tail occurs on the 3rd toss and C is the event that exactly two tails occurs in the 3 tosses, examine wheather events (i) A and B (ii) B and C are independent.

[G.U. (MA/MSc) '02]

Solution :

Total number of cases $= 2^3 = 8$

Favourable cases for event A are = HHT, HHH

Favourable cases for event B are = HHT, TTT, THT, HTT

Favourable cases for event C are = TTH, THT, HTT

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

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

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

(i) Favourable cases for $(A \cap B)$ is HHT

$$P(A \cap B) = \frac{1}{8}$$

$$P(A)\,P(B) = \frac{1}{4} \times \frac{1}{2} = \frac{1}{8}$$

$\therefore P(A \cap B) = P(A)\,P(A)$

$\therefore$ A and B are independent events.

(ii) Favourable cases for $(B \cap C)$ is HTT, THT

$$\therefore P(B \cap C) = \frac{2}{8} = \frac{1}{4}$$

$$P(B)\ P(C) = \frac{1}{2} \times \frac{3}{8} = \frac{3}{16}$$

Here, $P(B \cap C) \neq P(B)\ P(C)$

$\therefore$ B and C are not independent events.

Example 8. The odds that A speaks the truth are 3 : 2 and odds that B speaks the truth are 5 : 3. In what percentage of cases are they likely to contradict each other on an identical point?

[G.U. (MA/MSc) '03]

Solution : Given,

$$\text{Probability that A speaks truth } = \frac{3}{3+2} = \frac{3}{5}$$

$$\text{Probability that B speaks truth } = \frac{5}{5+3} = \frac{5}{8}$$

We have the following mutually exclusive events.

Let, a be the event that A speaks truth and B does not

b be the event that B speaks truth and A does not

$$P(a) = \frac{3}{5} \times \left(1 - \frac{5}{8}\right) = \frac{9}{40}$$

$$P(b) = \left(1 - \frac{3}{5}\right) \times \frac{5}{8} = \frac{1}{4}$$

$\therefore$ Required probability $P(a \cup b) = P(a) + P(b)$

$$\frac{9}{40} + \frac{1}{4} = \frac{19}{40}$$

$\therefore$ Percentage of cases in which they are likely to contradict each other is,

$$\frac{19}{40} \times 100\% = 47.5\%$$

Example 9. 12 balls are distributed at random among 3 boxes. What is the probability that the first box will contain 3 balls ?

[G.U. (MA/MSc) '04]

Solution :

Total number of cases $= 3^{12}$

From 12 balls 3 balls can be choosen in $^{12}C_3$ ways and remaining 9 balls can be distributed to the remaining two boxes in 2^9 ways.

∴ Favourable number of cases,

$$^{12}C_3 \times 2^9$$

$$\therefore \text{ Required probability } = \frac{^{12}C_3 \times 2^9}{3^{12}} = 0.212$$

CHECK YOUR PROGRESS

1. What are the basic probability rules?

[G.U. (MA/MSc) Prev. '05, 10]

2. Give the axiomatic definition of probability.
[GU(MA/MSc) Prev '06]

3. Define Baye's theorem. [GU(MA/MSc) Prev '05, '07, '09]

4. Define conditional probability. An urn contains 7 white, and 5 black balls. Two balls are drawn at random one after the other without replacement. Find the probability that both balls drawn are black.
[GU(MA/MSc) Prev '05, '10]

5. What are the basic theorems of probability? In a single throw with two dices, find the probability of getting a sum of (a) 8 and (b) 11.
[GU(MA/MSc) Prev '09]

6. Explain the concepts of equivalent events, independent events and conditional probability. Use the theorem of total probability to show that $P(A + B) = P(A\bar{B}) + P(\bar{A}B) + P(BA)$
[GU(MA/MSc) Prev '93]

7. In a class there are 40 boys and a certain number of girls. The probability of selecting a girl at random is 3/7. How many girls are there in the class? [GU(MA/MSc) '00]

8. If A and B are any two events in the sample space S which are not mutually exclusive events, show that (i) $P(A) \geq P(A \cap B)$ (ii) $P(A) \leq P(A \cup B)$ [GU(MA/MSc) Prev '98]

9. The probability that a boy will get a scholarship is 0.90 and that a girl win get it is 0.80. What is the probability that at least one of them will get the scholarship? [GU(MA/MSc) Prev '92]

1.7 Random Variable – Discrete and Continuous :

A random variable is a numerical valued function defined on a sample space. The variable is called random variable because its value is determined by the random outcome of the experiment i.e. the outcome which depends on chance only. If S is a sample space with a probability measure and X is a real valued function defined over the outcomes of S (denoted by X) then X is termed as a random variable. The set of values which X assumes is called the 'spectrum' of the random variable X.

For example, if we toss three unbiased coins, the number of head (X), constitute a discrete random variable. In notation of set theory, X can be expressed as,

$X = (x : x = 0, 1, 2, 3)$

In this example, X is a discrete random variable. More precisety, if X is a real valued function defined on a discrete random space, it is called a discrete random variable. In other words, if the random variable assumes only a finite or countably infinite set of values it is known as discrete random variable.

On the other hand, if X is a real valued function defined over a continuous sample space, it is called a continuous random variable. In other words, if the random variable can assume all possible values between certain limits i.e. uncountably infinite number of values, it is known as continuous random variable. For example, if X denotes the height of students of the Economics Department of Gauhati University, X is a continuous random variable, for it can assume any value, say between 5 feet and 6 feet. In set notation,

$X = \{x : 5 \leq x \leq 6\}$

1.7.1 Probability Mass Function :

If X is a discrete random variable taking at most a countably infinite number of values x_1, x_2 .... then its probabilistic behavior at each real point is described by a function which is called probability mass function. It is defined below :

If X is a discrete random variable, the function given by $p(x) = P(X = x)$ for each particular value, x assumed by X within its elements, is called the probability mass function (p.m.f) of X.

A function p(x) can serve as the probability mass function of a discrete random variable X, if it satisfies the following conditions.

1. $p(x) \geq 0$ for each value within the domain of X.

2. $\sum_x p(x) = 1$,

(The summation extending over all the values assumed by X)

1.7.2 Probability destribution of a Discrete Random variable X :

The set of all possible ordered pairs of values assumed by the discrete random variable X and the corresponding probabilities, $\{x, p(x)\}$ is termed as the probability distribution of the discrete random variable X. It tells us how the total probability of a discrete random variable X is distributed among the different values assumed by X. It is usually depicted in a tabular form as follows, Probability distribution of the Discrete Random Variable X.

$X = x$:	x_1	x_2	......	x_n
$p(X = x)$:	$p(x_1) = p_1$	$p(x_2) = p_2$	....	$p(x_n) = p_n$

Example,

If three unbiased coins are thrown, then the probability distribution of the number of heads (X) is as follows,

$X = x$	0	1	2	3
$p(X = x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Note :

Random variables are also called simply variates and are denoted by capital letters X, Y, ... whereas their specific values are denoted respectively by small letters x, y, ...

1.7.3 Probability Density Function :

If X is a continuous random variable, the probability that X assumes a particular value, say c, is effectively zero, that is $P(X = c) = 0$. In such a case, it is more relevant to obtain the probability that the continuous random variable X lies within a range of values say between a and b where $a < b$.

The definiton of probability in the continuous case presumes, for such a variable, the existence of a function, called the probability density function (p.d.f) such that by integrating it within some specified limit we obtain the probability that the variable X lies within that particular range of values.

In other words, if f(x) is the p.d.f of the continuous random variable X, then

$$p(a \le x \le b) = \int_a^b f(x)dx$$

Where a and b are any real constants, with $a \le b$

A function f(x) can serve as the p.d.f of a continuous random variable X which can assume any value between $-\infty$ and $+\infty$, only if it satisfies the following conditions

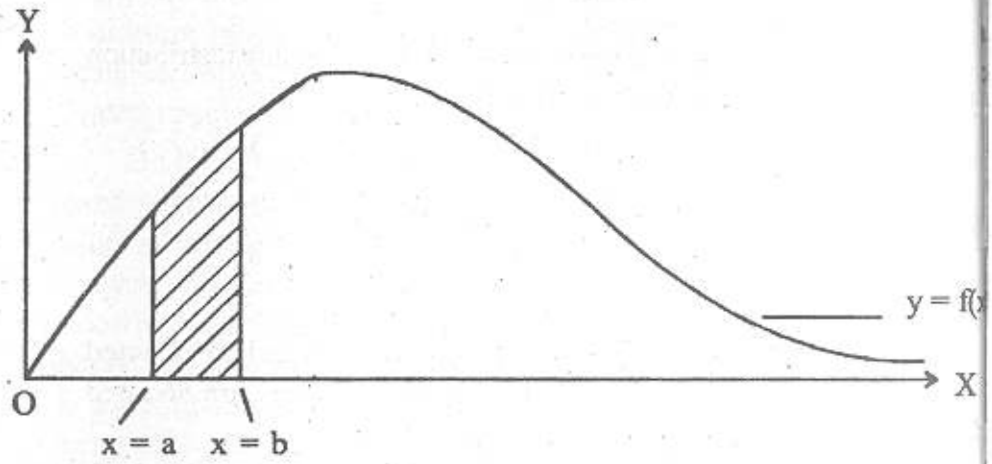
(i) $f(x) \ge 0$, for $-\infty \le x \le \infty$

(ii) $\int_{-\infty}^{\infty} f(x)dx = 1$

(iii) The probability P(E) given by

$$P(E) = \int_E f(x)dx$$

is well defined for any event E.



In the diagram, the probability that the continuous random variab X lies between a and b is given by the area bounded by the y = f(curve, the X axis and the ordinates at the points X = a and X = b. T y = f(x) curve is known as the probability density curve or simply t probability curve. The expression f(x) dx, also written as d F(x) is know as the probability differential.

In case of discrete random variable, the probability at a point P(X = c) is not zero for some fixed c. However, in case of continuo random variables the probability at a point is always zero i. (P(X = c) = 0) for all possible values of c.

SELF-ASKING QUESTION

Try yourself to distinguish between probability mass function an probability density function

1.8 Mathematical Expectation :

If X is a random variable which can assume any one of the valu $x_1, x_2, \ldots, x_n$ with respective probabilities $p_1, p_2, \ldots, p_n$ then t mathematical expectation of X, usually called the expected value of and denoted by E(x) is defined as,

$$E(X) = p_1x_1 + p_2x_2 + \ldots + p_nx_n$$

$$= \Sigma p_ix_i \qquad \text{where, } p_i = p_1 + p_2 + \ldots + p_n = 1$$

If X is a discrete random variable and f(x) is the value of its p.m.f. at the point X = x, the mathematical expectation or expected value of the discrete random variable X is given by

$$E(X) = \sum_{x} x\, f(x)$$

If X is a continuous random variable and f(x) is the value of its p.d.f. at X = x, the mathematical expectation or expected value of the continuous random variable X is given by,

$$E(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx$$

In the above two difinitious, it is assumed that the sum or the integral exists otherwise, the mathematical expectation ceases to exist.

Example :

4 coins are tossed. What is the mathematical expectation of obtaining heads ?

Solution :

Let, the descrete random variable X denotes the number of heads obtained when 4 coins are tossed. We get the following p.d.f of X.

X	0	1	2	3	4
p(X)	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$

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

$$= 0 \times \frac{1}{16} + 1 \times \frac{4}{16} + 2 \times \frac{6}{16} + 3 \times \frac{4}{16} + 4 \times \frac{1}{16} = \frac{32}{16} = 2$$

In the above example we cannot expect to obtain the value 2 in actual experiment for all times. Rather, if an experiment is repeated a large number of times under essentially homogenous conditions, then the average or mean of the actual outcomes constitute the expected value of the variable.

STOP TO CONSIDER

Theorems on Mathematical Expectation :

(i) If 'a' is a constant, then E(a) = a.

Proof :

Let X be the random variable which takes the values $x_1, x_2, ..., x_n$

with respective probabilities

$P(x_1), P(x_2), \ldots, P(x_n)$

$$E(X) = \sum_{i=1}^{n} x_i P(x_i) = x_1 P(x_1) + x_2 P(x_2) + \ldots + x_n P(x_n)$$

Let, X = a (constant)

$$\begin{aligned} E(a) &= a\,P(x_1) + a\,P(x_2) + \ldots + a\,P(x_n) \\ &= a\,\{P(x_1) + P(x_2) + \ldots + P(x_n)\} \\ &= a.\,1 \\ &= a \end{aligned}$$

(ii) If 'a' is constant and 'X' is a random variable then

$$E(aX) = a\,E(X)$$

Proof :

For a random variable X, we have

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

$$\therefore E(aX) = \sum_{i=1}^{n} a x_i p_i = a \sum_{i=1}^{n} x_i p_i = a\,E(X)$$

(iii) E {X – E(X)} = 0

Proof :

$$\begin{aligned} E\{X - E(X)\} &= E(X) - E\{E(X)\} \\ &= E(X) - E(X) \qquad [\because E(X) = \text{constant}] \\ &= 0 \end{aligned}$$

(iv) If a and b are constants and X is a random variable, then

$$E(aX + b) = aE(X) + b$$

Proof :

Let X be a random variable which takes the values $x_1, x_2, \ldots, x_n$ with respective probabilities $p_1, p_2, \ldots, p_n$.

$$\begin{aligned} E\,(aX + b) &= \sum_{i=1}^{n} (a x_i + b)\, p_i \\ &= (ax_1 + b)p_1 + (ax_2 + b)p_2 + \ldots + (ax_n + b)\,P_n \\ &= a(x_1 p_1 + x_2 p_2 + \ldots + x_n p_n) + b(p_1 + p_2 + \ldots + p_n) \\ &= a \sum_{i=1}^{n} x_i p_i + b \sum_{i=1}^{n} p_i \\ &= a\,E(X) + b.1 \\ &= a\,E(X) + b. \end{aligned}$$

1.8.1 Rules of Mathematical Expectation :

(a) The Additive Rule of Expectation :

The mathematical expectation of the sum of random variables is equal to the sum of their expectations, provided all the expectations exist.

Thus, if X and Y are two random variables, then

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

Proof :

Suppose, the random variable X assumes the values $x_1, x_2, ..., x_n$ with respective probabilities $p_1, p_2, ..., p_n$ then expectation is

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

Let, Y be the other random variable, which assumes values $y_1, y_2, ..., y_m$ with respective probabilities $p'_1, p'_2, ..., p'_m$. Then the expectation is,

$$E(Y) = \sum_{j=1}^{m} y_j p'_j ....(ii)$$

Again, the sum (X + Y) would be a random variable which can take $m \times n$ values $(x_i + y_j)$ $i = 1, 2, ... n, j = 1, 2, ..., m.$

By definition,

$$E(X + Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} (x_i + y_j) p_{ij}$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{m} x_i p_{ij} + \sum_{i=1}^{n} \sum_{j=1}^{m} y_j p_{ij}$$

$$= \sum_{i=1}^{n} x_i \sum_{j=1}^{m} p_{ij} + \sum_{j=1}^{m} y_j \sum_{i=1}^{n} p_{ij}$$

$$= \sum_{i=1}^{n} x_i p_i + \sum_{j=1}^{m} y_j p_j^{\prime}$$

$$= E(X) + E(Y) \qquad \text{(using (i) and (ii))}$$

The above result can be extended to any number of variables. Thus, if $A_1, A_2, ... A_n$ are n random variables, we have,

$$E(A_1 + A_2 + ... + A_n) = E(A_1) + E(A_2) + ... E(A_n)$$

In the above theorem if X and Y assume a finite number of values, then E(X) and E(Y) always exist. However, if X and Y assume an infinite number of values, then the additive law of expectation viz $E(X + Y) = E(X) + E(Y)$ holds only if the exputations exist, that is provided

$$\sum_{i=1}^{\infty} |x_i p_i| < \infty \qquad \text{and} \qquad \sum_{j=1}^{\infty} |y_j p_j^{\prime}| < \infty$$

(b) Expectation of a Linear combination of Random Variable

If $x_1, x_2, \ldots, x_n$ are any n random variables and if $a_1, a_2, \ldots, a_n$ are any n constants, then

$$E\left(\sum_{i=1}^{n} a_i x_i\right) = \sum_{i=1}^{n} a_i E(x_i)$$

(c) The Multiplicative Law of Expectation :

The mathematical expectation of the product of a number of independent random variables is equal to the product of their expectations.

Thus, for any two independent random variables X and Y.

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

Proof :

Suppose, X is an independent random variable which takes the values $x_1, x_2, \ldots, x_n$ with respective probabilities $p_1, p_2, \ldots p_n$, then mathematical expectations is,

$$E(X) = \sum_{i=1}^{n} x_i p_i \text{ .... (i)}$$

Let, Y be the another independent random variable, which can take the values $y_1, y_2, \ldots, y_m$ with respective probabilities $p_1', p_2', \ldots, p_m'$.

The mathematical expectation is,

$$E(Y) = \sum_{j=1}^{m} y_j p_j' \text{ ....(ii)}$$

The product XY would also be a random variable which can assume mn values $x_i y_i$ $(i = 1, 2, \ldots n)$ $(j = 1, 2, \ldots, m)$

By definition,

$$E(XY) = \sum_{i=1}^{n} \sum_{j=1}^{m} x_i y_j p_i p_j'$$

$$= \sum_{i=1}^{n} x_i p_i \sum_{j=1}^{m} y_j p_j'$$

$$= E(X)\, E(Y) \qquad \text{(Using (i) and (ii))}$$

The above result can be extended to any number of independent random variables.

It should be noted that the multiplicative theorem of expectation holds only for independent events while no such condition on the variables is required for the additive theorem of expectation.

SELF-ASKING QUESTION

Try yourself to prove the rules of expectation with the help of simple numerical example.

1.8.2 Variance in terms of Expectation

If X is a random variable, then $Var(X) = E(X^2) - \{E(X)\}^2$

$$\begin{aligned} Var(X) &= E[X - E(X)]^2 \\ &= E[X^2 - 2XE(X) + \{E(X)\}^2] \\ &= E(X^2) - 2\{E(X)\}^2 + \{E(X)]^2 \\ &= E(X^2) - \{E(X)\}^2 \end{aligned}$$

Theorems on variance :

(i) Var (C) = 0 where C is a constant.

Proof :

We have,

$$\begin{aligned} Var(X) &= E[X - E(X)]^2 \\ Var(C) &= E[C - E(C)]^2 \\ &= E[C - C]^2 \\ &= E(0)^2 = 0 \end{aligned}$$

(ii) Var (X ± C) = Var (X)

Thus, variance is independent of change of origin.

Proof :

We have,

$$\begin{aligned} E(X + C) &= E(X) + E(C) \\ &= E(X) + C \end{aligned}$$

By definition,

$$\begin{aligned} Var(X + C) &= E[(X + C) - E(X - C)]^2 \\ &= E[X - E(X)]^2 \\ &= Var X. \end{aligned}$$

Similarly, it can be proved that,

$$Var(X - C) = Var(X)$$

(iii) Var (aX) = a^2 Var (X)

Thus, variance is not independent of change of scale.

Proof :

$$\begin{aligned} Var(aX) &= E[aX - E(aX)]^2 \\ &= E[aX - aE(X)]^2 \\ &= E[a\{X - E(X)\}]^2 \\ &= a^2 E[X - E(X)]^2 \\ &= a^2 Var X. \end{aligned}$$

1.8.3 Covariance in terms of Expectation :

If X and Y are two variables with respective values (or means) $\overline{X}$ and $\overline{Y}$, then the covariance between X and Y is defined as,

$$\begin{aligned} Cov(X, Y) &= E[(X - \overline{X})(Y - \overline{Y})] \\ &= E[XY - XE(Y) - E(X)Y + E(X)E(Y)] \\ &\qquad [\because \overline{X} = E(X) \ \& \ \overline{Y} = E(Y)] \\ &= E(XY) - E(X)E(Y) - E(X)E(Y) + E(X)E(Y) \\ &= E(XY) - E(X)E(Y) \end{aligned}$$

If X and Y are independent,

$$Cov(X,Y) = E\left[\left(X-\overline{X}\right)\left(Y-\overline{Y}\right)\right] = 0$$

Thus, the covariance of two independent variables is zero.

STOP TO CONSIDER

Variance of a Linear Combination of Random Variables :

Let $X_1, X_2, \ldots, X_n$ be n random variables with finite variances $\sigma_1^2\ \sigma_2^2, \ldots, \sigma_n^2$.

Thus, $t = a_1X_1 + a_2X_2 + \ldots + a_nX_n$ is a linear combination of random variables where $a_1, a_2, \ldots, a_n$ are constant.

Now, $E(t) = a_1E(X_1) + a_2E(X_2) + \ldots + a_nE(X_n)$

$$\Rightarrow t - E(t) = a_1[X_1 - E(X_1)] + a_2[X_2 - E(X_2)] + \ldots + a_n[X_n - E(X_n)]$$

Squaring both sides and taking the expected values we have,

$Var(t) = a_1^2 Var(X_1) + a_2^2 Var(X_2) + \ldots + a_n^2 Var(X_n) + 2a_1a_2 Cov(X_1, X_2) + \ldots + 2a_{n-1}a_n Cov(X_{n-1}, X_n)$

$$\sum_{i=1}^{n} a_i^2 Var(X_i) + 2\sum_{i<j=1}^{n} a_ia_jCov(X_i, X_j)$$

Corollary I :

If $a_1 = 1, a_2 = 1, a_3 = a_4 = \ldots = a_n = 0$

then $Var(X_1 + X_2) = Var(X_1) + Var(X_2) + 2\,Cov(X_1, X_2)$

Corollary II :

If $a_1 = 1, a_2 = -1, a_3 = a_4 = \ldots = a_n = 0$

then $Var(X_1 - X_2) = Var(X_1) + Var(X_2) - 2\,Cov(X_1, X_2)$

Corollary III :

If X_1 and X_2 are independent variables, then Cov $(X_1, X_2) = 0$

Hence, Var $(X_1 \pm X_2)$ = Var (X_1) + Var (X_2)

1.9 Solved Examples :

Example 1. A box contains 'a' white balls, 'b' black balls, 'c' balls are drawn. Find the expectations of the number of white balls drawn.

[GU (MA/MSc) Prev. '92]

Solution :

Let us associate a variable X_i with the i-th draw such that,

$x_i = 1$ if the i-th draw results in a white ball.

$x_i = 0$ if the i-th draw results in a black ball.

Then the number of white balls drawn among c balls is

$$S = x_1 + x_2 + ... + x_c$$

$$\therefore E(S) = E(x_1 + x_2 + ... + x_c)$$

$$= E(x_1) + E(x_2) + ... + E(x_c)$$

Now, $E(x_i) = 1 \ . \ p(x_i = 1) + 0 \ . \ p \ (x_i = 0)$

$$= 1 . \frac{a}{a+b} + 0 . \frac{b}{a+b}$$

$$= \frac{a}{a+b} \qquad (i = 1, 2, ... c)$$

$$\therefore E(S) = \frac{a}{a+b} + \frac{a}{a+b} + ... + \frac{a}{a+b} \quad \text{c times}$$

$$= c \frac{a}{a+b}$$

Example 2. A box contains 12 items of which 3 are defective. A sample of 3 items is selected at random from this box. If x denotes the number of defective items, find mathematical expectation and variance of X.

[GU (MA/MSc) Prev. '94]

Solution : 3 balls can be drawn out of 12 in ${}^{12}C_3$ ways.

If X denotes the number of defective items we have the following probability distribution.

X :	0	1	2	3
P(X) :	$84/220$	$108/220$	$27/220$	$1/220$

How we have got the table values ? Like this,

$$P(X=0)=\frac{{}^9C_3}{{}^{12}C_3}=\frac{84}{220}, \qquad P(X=1)=\frac{{}^9C_2\times{}^3C_1}{{}^{12}C_3}=\frac{108}{220}$$

$$P(X=2)=\frac{{}^9C_1\times{}^3C_2}{{}^{12}C_3}=\frac{27}{220}$$

$$P(X=3)=\frac{{}^9C_0}{{}^{12}C_3}=\frac{1}{220}$$

Hint : In this case, $P(X=0)$ implies all balls drawn are not defect Similarly, $P(X=1)$ implies out of 3 balls 1 is defective and the oth are not defective.

$$\text{So, } E(X)=\sum_{i=1}^{n} x_i P(x_i)$$

$$= 0\times {}^{84}/_{220}+1\times {}^{108}/_{220}+2\times {}^{27}/_{220}+3\times {}^{1}/_{220}$$

$$=\frac{165}{220}$$

We know that,

$$\text{Var }(X)=E(X^2)-\{E(X)\}^2$$

$$\text{Now, } E(X^2)=0^2\times {}^{84}/_{220}+1^2\times {}^{108}/_{220}+2^2\times {}^{27}/_{220}+3^2\times {}^{1}/_{22}$$

$$=\frac{225}{220}$$

$$\text{Var }(X)=\frac{225}{220}-\left(\frac{165}{220}\right)^2$$

$$=\frac{49500-27225}{48400}=0.46$$

Example 3. Suppose that we have a series of n independent trials, t probability of success being p_i for the i-th trial. Show that the me number of success is $\sum_{i=1}^{n} p_i$. Also find the variance of number of succe

[G.U (MA/MSc) Prev. '9

Solution :

Let, $X_i=1$ if there is success in *i*th trial and $X_i=0$ if there is failu in *i*th trial.

Thus, we get the following probability distribution of X_i.

X_i : 0 1

$P(X_i)$: $(1 - p_i)$ p_i

$$\therefore E(X_i) = \sum_{i=1}^{n} x_i p_i$$
$$= 0 \times (1 - p_i) + 1 \times p_i$$
$$= p_i$$

Let, S is the number of success in n trials

$$S = X_1 + X_2 + ... + X_n$$
$$\therefore E(S) = E(X_1) + E(X_2) + ... + E(X_n)$$
$$= p_1 + p_2 + ... + p_n = \sum_{i=1}^{n} p_i$$

Variance of $X_i = E(X_i^2) - \{E(X_i)\}^2$

$$E(X_i^2) = \sum_{i=1}^{n} x_i^2 p_i = 0^2.(1 - p_i) + 1^2.p_i = p_i$$

$$Var(X_i) = p_i - p_i^2$$
$$\therefore Var(S) = Var(X_1) + Var(X_2) + ... + Var(X_n)$$
$$= (p_1 - p_1^2) + (p_2 - p_2^2) + ... + (p_n - p_n^2)$$
$$= \sum_{i=1}^{n} (p_i - p_i^2)$$

Example 4. 3 coins whose faces are marked 1 and 2, are tossed. What is the expectation of the total value of numbers on their faces ? [G.U. (MA/MSc) Prev. '98]

Solution :

Let, X denotes the total value of numbers of 3 coins. We have the following probability distribution of x.

X :	3	4	5	6
P(X) :	$1/8$	$3/8$	$3/8$	$1/8$

[Try to find out how P(X) s are determined]

$$E(X) = \sum_{i=1}^{n} x_i p_i = 3 \times \frac{1}{8} + 4 \times \frac{3}{8} + 5 \times \frac{3}{8} + 6 \times \frac{1}{8} = \frac{36}{8} = 4.5$$

Example 5. If two dice are thrown find the mathematical expectation of the sum of the numbers. [G.U. (MA/MSc) '00]

Solution :

Let, X denotes the sum of the numbers when 2 dice are thrown. We get the following probability distribution.

X :	2	3	4	5	6	7	8
P(X) :	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$
X :	9	10	11	12			
P(X) :	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$			

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

$$= 2\times\frac{1}{36}+3\times\frac{2}{36}+4\times\frac{3}{36}+5\times\frac{4}{36}+6\times\frac{5}{36}+$$

$$7\times\frac{6}{36}+8\times\frac{5}{36}+9\times\frac{4}{36}+10\times\frac{3}{36}+11\times\frac{2}{36}+12\times\frac{1}{36}$$

$$= \frac{252}{36} = 7$$

Example 6. If 'n' dice are thrown, find the expectation of the

(i) sum of points on 'n' dice.

(ii) Product of the points on 'n' dice [G.U. (MA/MSc) Prev. '0

Solution :

Let, X_i denotes the number on a die when thrown.

Then X can take any one of the values 1, 2, 3, 4, 5, 6 each wi

equal probability $\frac{1}{6}$.

Hence, $E(X_i) \quad = \frac{1}{6}\times 1+\frac{1}{6}\times 2+....+\frac{1}{6}\times 6 = \frac{7}{2} \qquad (i = 1, 2, ...$

(i) Now, the sum of points obtained on 'n' die, S is given by

$$S = X_1 + X_2 + ... + X_n$$

$$\therefore E(S) = E(X_1 + X_2 + ... + X_n)$$

$$= E(X_1) + E(X_2) + ... + E(X_n)$$

$$= \frac{7}{2}+\frac{7}{2}+\frac{7}{2}+...+\frac{7}{2} \quad (n \text{ times})$$

Hence $E(S) = \frac{7n}{2}$

(ii) The product of the points obtained on 'n' dice is

$$P = X_1 . X_2 .... X_n$$

$\therefore E(P) = E(X_1 . X_2 ..... X_n)$

$= E(X_1)\, E(X_2) ... E(X_n)$

(Since $X_1, X_2, ..., X_n$ are independent)

$= \frac{7}{2} . \frac{7}{2} .... \text{n times}$

$= \left(\frac{7}{2}\right)^n$

Example 7. A coin is tossed until a head appears. What is the expectation of the number of tosses required? **[G.U. (MA/MSc) '01]**

Solution :

Let, X denotes the number of tosses required. We get the following probability distribution of X.

Now,

Event	X	P(X)
H	1	$\frac{1}{2}$
TH	2	$\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$
TTH	3	$\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}$
TTTH	4	$\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{16}$
⋮	⋮	⋮

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

$= 1 \times \frac{1}{2} + 2 \times \frac{1}{4} + 3 \times \frac{1}{8} + 4 \times \frac{1}{16} + ...$

Let $E(X) = S$

$\therefore\ S = 1 \times 1/2 + 2 \times 1/4 + 3 \times 1/8 + 4 \times 1/16 + ... \qquad (i)$

$\Rightarrow \frac{1}{2} S = \frac{1}{4} + 2 \times \frac{1}{8} + 3 \times \frac{1}{16} + 4 \times \frac{1}{32} + ... \qquad (ii)$

$$\text{(i)} - \text{(ii)} \Rightarrow S - \frac{1}{2}S = \left(\frac{1}{2} + 2\times\frac{1}{4} + 3\times\frac{1}{8} + 4\times\frac{1}{16} + \ldots\ldots\right)$$

$$-\left(\frac{1}{4} + 2\times\frac{1}{8} + 3\times\frac{1}{16} + 4\times\frac{1}{32} + \ldots\right)$$

$$\Rightarrow \frac{2S - S}{2} = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \frac{1}{32} + \ldots$$

$$\Rightarrow \frac{S}{2} = \tfrac{1}{2} + \tfrac{1}{2}\times\left(\tfrac{1}{2}\right) + \tfrac{1}{2}\left(\tfrac{1}{2}\right)^2 + \tfrac{1}{2}\left(\tfrac{1}{2}\right)^3 + \tfrac{1}{2}\left(\tfrac{1}{2}\right)^4 + \ldots$$

(Which is a infinity term of G.P. Series)

$$\Rightarrow \frac{S}{2} = \frac{\tfrac{1}{2}}{1 - \tfrac{1}{2}}$$

$\Rightarrow S = 2$

$\therefore\ E(X) = 2$

Expected number of tosses is 2.

Example 8. A man wins a rupee for a head and loss a rupee for a tail, when a coin is tossed. Suppose, that he tosses once and quits if he wins, but tries once more if he losses on the first toss. What is his expected gain? [G.U (MA/MSc) '04]

Solution :

Let, X denotes the gain of the man in rupees. Thus, we get the following probability distribution function of X.

Event :	H	TH	TT
X :	1	$-1+1=0$	$-1-1=-2$
P(X) :	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$

$$E(X) = \sum_{i=1}^{n} x_i p(x_i) = 1\times\frac{1}{2} + 0\times\frac{1}{4} - 2\times\frac{1}{4} = 0$$

Thus, the man has neither expected gain nor expected loss.

Example 9. A die is tossed twice. Obtaining a number less than 3 is termed as a success, obtain the probability distribution and hence the mathematical expectation. [G.U. (MA/MSc) '05]

Solution :

Let, X denotes the number of success. We have the following probability distribution function of X.

X :	0	1	2
P(X) :	$\frac{4}{9}$	$\frac{4}{9}$	$\frac{1}{9}$

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

$$= 0\times\frac{4}{9}+1\times\frac{4}{9}+2\times\frac{4}{9}$$

$$= 1.33$$

CHECK YOUR PROGRESS

1. Define mathematical expectation of a random variable. State the basic theorems of mathematical expectation. Prove that E(X + Y) = E(X) + E(Y) [GU(MA/MSc) Prev '05]
2. Define mathematical expectation of a random variable. If X and Y are independent events, show that E(XY) = E(X) . E(Y). [GU(MA/MSc) Prev '06]
3. Define Random variable. [GU(MA/MSc) Prev '08]
4. Give various theorems of mathematical expectation. [GU(MA/MSc) Prev '09]
5. Show that Var (X) = $E(X^2) - \{E(X)\}^2$ [GU(MA/MSc) Prev '91, '96]
6. Show that the covariance of two independent variables in equal to zero. But the converse of this is not necessarily true. [GU(MA/MSc) Prev '02, '03]
7. Find the expectation and variance of the number of successes in a series of 'n' independent trials. The probability of success in the i-th trial being P_i. [GU(MA/MSc) Prev '06]
8. Two cards are drawn without replacement from a well shuffled pack of cards. Obtain the probability distribution and hence the expected number of face cards (Jack, Queen, King and Ace).
9. Obtain the mean and variance of $Y = 2x_1 + 2x_2 + 4x_3$ where x_1, x_2 and x_3 are three random variables with means given by 3, 4, 5 respectively and variance by 10, 20 and 30 respectively, and covariance by $\sigma_{12} = 0$, $\sigma_{23} = 0$, $\sigma_{13} = 5$ where $\sigma_{ij} = \text{Cov}(x_i, x_j)$

1.10 Moments : In Statistics, we talk of moments of random variable about some point. The term moment is used to describe the various characteristics of a frequency distribution like central tendency, variation, skewness and kurtosis etc.

1.10.1 Moments about the Origin :

If X is a discrete random variable, the rth moment about the origi of the distribution of X, is the expected value of X^r.

In symbol,

$\mu_r' = E(X^r) = \sum x^r p(x)$ $\quad$ $r = 0, 1, 2, 3 \ldots$ (i)

where p(x) is the p.m.f at X = x

If X is a continuous random variable,

$$\mu_r' = E(X^r) = \int_{-\infty}^{\infty} x^r f(x)dx, \quad \text{where f(x) is the p.d.f of X.}$$

From equation (i) it follows that,

$\mu_0' = E(X^0) = E(1) = 1$

$\mu_1' = E(X^1) = \mu$ (the mean of the distribution)

Hence the first moment about the origin is the mean of the distribution.

$$\mu_2' = E(X^2) = \sum_x x^2 p(x)$$

$$\mu_3' = E(X^3) = \sum_x x^3 p(x)$$

1.10.2 Moments about the mean or Central Moments :

If X is a discrete random variable, its rth moment about the mean, termed as the rth central moment, denoted by μ_r is the expected value of $(X - \mu)^r$.

In symbol,

$\mu_r = E[(X - \mu)^r] = \sum_x (x - \mu)^r p(x),$ $\quad$ $r = 0, 1, 2, 3 \ldots$ (2)

If X is a continuous random variable.

$$\mu_r = E[(X - \mu)^r] = \int_{-\infty}^{\infty} (x - \mu)^r f(x)dx$$

From equation (2) it follows that,

$\mu_0 = E[(X - \mu)^0] = E(1) = 1$

$\mu_1 = E[(X - \mu)^1] = E(X) - E(\mu) = 0$ $\quad$ [since $E(X) = E(\mu)$]

$\mu_2 = E[(X - \mu)^2]$

The second central moment is the variance of the distribution of X and shows the disperssion about the mean. It is conventionally denoted by σ^2.

Central Moments expressed in terms of Moments about the origin :

$$\mu_2 = E[(X-\mu)^2]$$
$$= E(X^2 - 2\mu X + \mu^2)$$
$$= E(X^2) - 2\mu E(X) + E(\mu^2)$$
$$= E(X^2) - \mu^2 \qquad [\text{since } E(X) = \mu = \mu_1']$$
$$= \mu_2' - \mu_1'^2$$

Similarly,

$$\mu_3 = E[(X-\mu)^3]$$
$$= E(X^3 - 3\mu X^2 + 3\mu^2 X - \mu^3)$$
$$= E(X^3) - 3\mu E(X^2) + 3\mu^2 E(X) - E(\mu^3)$$
$$= \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3 \qquad [\text{since } E(X) = \mu = \mu_1']$$

In the same way, it can be shown that

$$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4$$

In general,

$$\mu_r = E[(X-\mu)^r]$$
$$= E[X^r - {}^rC_1 X^{r-1}\mu + {}^rC_2 X^{r-2}\mu^2 \ldots + (-1)^r \mu^r]$$
$$= \mu_r' - {}^rC_1 \mu_{r-1}' \mu_1' + {}^rC_2 \mu_{r-2}' \mu_1'^2 \ldots\ldots + (-1)^r \mu_1'^r$$

Putting r = 2, 3, 4, ... in the above equation, μ_2, μ_3, μ_4 ... can be obtained in terms of moments about the origin.

1.10.3 Moments about an Arbitrary Constant (a) :

The r-th moment about an arbitrary constant 'a' is defined as follows,

$$\mu_r^a = E(X-a)^r$$

Thus, $\mu_0^a = E(X-a)^0 = E(1) = 1$

$$\mu_1^a = E(X-a)^1 = E(X) - E(a) = \mu_1' - a$$
$$\mu_2^a = E(X-a)^2 = \mu_2' - 2a\mu_1' + a^2$$
$$\mu_3^a = E(X-a)^3 = \mu_3' - 3a\mu_2' + 3a^2\mu_1' - a^3$$
$$\mu_4^a = E(x-a)^4 = \mu_4' - 4a\mu_3' + 6a^2\mu_2' - 4a^3\mu_1' - a^4$$

In general,

$$\mu_r^a = E(X-a)^r$$
$$= E[X^r - {}^rC_1 X^{r-1}a + {}^rC_2 X^{r-2}a^2 - \ldots\ldots + (-1)^r a^r]$$
$$= \mu_r' - {}^rC_1 a\mu_{r-1}' + {}^rC_2 a^2 \mu_{r-2}' - \ldots + (-1)^r a^r.$$

Thus, if the moments about any arbitrary points are given, the corresponding moments about the origin can be obtained by using the above relations.

Moments about the origin and moments about any arbitrary constant 'a' are also known as raw moments.

1.10.4 Moments about the mean in terms of moments about a arbitrary constant a :

$\mu_r = E(X - \mu)^2$

$= E[(X - a) + (a - \mu)]^r$

$= E[(X - a) - \mu_1^a]^r$

$= E[(X - a)^r - r_{C_1} (X - a)^{r-1} \mu_1^a + r_{C_2} (X - a)^{r-2} \mu_1^{a^2} + ... + (-1)^r \mu$

$= \mu_r^a - r_{C_1} \mu_{r-1}^a \mu_1^a + r_{C_2} \mu_{r-2}^a \mu_1^{a^2} + ... + (-1)^r \mu_1^{a^r}$

In particular, putting r = 2

$\mu_2 = \mu_2^a - (\mu_1^a)^2$

Similarly, putting r = 3, 4 we have,

$\mu_3 = \mu_3^a - 3\mu^a{}_2\mu_1^a + 2(\mu_1^a)^3$

$\mu_4 = \mu_4^a - 4\mu_3^a\mu_1^a + 6\mu_2^a\mu_1^{a^2} - 3\mu_1^{a^4}$

1.10.5 Moments about any Arbitrary constant 'a' in terms moments about the mean :

$\mu_r^a = E(X - a)^r$

$= E[(X - \mu) + (\mu - a)]^r$

$= E[(X - \mu) + \mu_1^a]^r$

$= E[(X - \mu)^r + r_{C_1} (X - \mu)^{r-1} \mu_1^a + r_{C_2} (X - \mu)^{r-2} \mu_1^{a^2} + ... + \mu$

$= \mu_r + r_{C_1} \mu_{r-1} \mu_1^a + r_{C_2} \mu_{r-2} \mu_1^{a^2} + ... + \mu_1^{a^r}$

In particular, putting r = 2, 3 and 4 in the above equation, we ha

$\mu_2^a = \mu_2 + (\mu_1^a)^2$

$\mu_3^a = \mu_3 + 3\mu_2\mu_1^a + (\mu_1^a)^3$

$\mu_4^a = \mu_4 + 4\mu_3\mu_1^a + 6\mu_2 (\mu_1^a)^2 + (\mu_1^a)^4$

STOP TO CONSIDER

Corollary :

1. For a symmetrical distribution, all the odd order central momen are equal to zero.
2. The central moments are invariant to the change of origin but n of scale.
3. In order to know about the nature and form of a particular distribution, we must have knowledge about a measure of cent

tendency ($\mu = \mu_1'$) a measure of disperssion ($\sigma^2 = \mu_2$) as well as measure of skewness and kurtosis.

For deriving the latter 2 measures, the first four central moments need to be estimated. Hence higher order moments are generally not estimated.Karl Pearson introduced the concept of Beta (β) and Gamma (γ) coeffecients, based on central moments, as a measure of skewness and kurtosis.

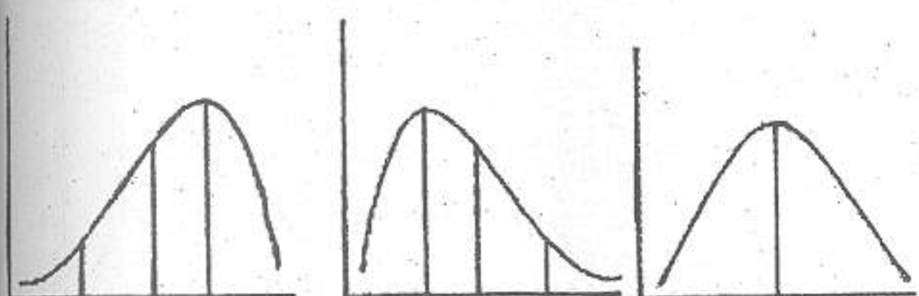
1.11 Skewness and Kurtosis :

Coeffecient of Skewness : Skewness refers to lack of symmetry i.e. when a distribution is not symmetrical, it is called a skewed distribution. A measure of skewness is obtained by making use of the second and third moments about mean. Here, β_1 is used as the coefficient (measure) of skewness.

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3}$$

$$\gamma_1 = +\sqrt{\beta_1} = \sqrt{\frac{\mu_3^2}{\mu_2^3}} = \frac{\mu_3}{\mu_2^{\frac{3}{2}}} = \frac{\mu_3}{(\sigma^2)^{3/2}} = \frac{\mu_3}{\sigma^3}$$

Skewness means, mean $\neq$ median $\neq$ mode.



Mean Medⁿ Mode | Mode Medⁿ Mean | Mean=Medⁿ=Mode

Mean<Med<Mode — *Negatively Skewed*

Mode<Med<Mean — *Positively skewed*

Mean = Median = Mode — *Symmetric*

Symmetrical curves represent normal destribution. The characteristics of these curves are that on being folded vertically from the middle, the two sides exactly coincide.

If $\beta_1 = 0$, the distribution is perfectly symmetrical, otherwise it is skewed. However, since β_1 is always positive it does not shed any light as to whether a particular distribution is positively skewed or negatively skewed, in the event of it being skewed ($\beta_1 \neq 0$). This limitation is removed in the case of Gamma coefficient γ_1, which is defined as positive square root of β_1. From the γ_1 coeffecient, it is evident that if is negative, γ_1 is negative and the distribution is negatively skewed. the otherhand, when μ_3 is positive, γ_1 is positive and the distributio positively skewed.

Let us understand,

- In case the distribution has a tail towards larger values we ha a positively skewed distribution.

- If the distribution has a tail towards smaller values we hav negatively skewed distribution.

- When Mean = Median = Mode (i.e. central values coincid there is symmetry.

STOP TO CONSIDER

Difference between Skewness and Dispersion :

Dispersion is concerned with the amount of variations rathe than with its directions. Skewness tells us about the direction of th variation or the departure from symmetry. In fact, measures of skewnes are dependent upon the amount of dispersion.

P.S. It may be noted that although skewness is an importan characteristic for defining the precise pattern of a distribution, it i rarely calculated in business and economic series. Variation is by fa the most important characteristic of a distribution.

Coefficient of Kurtosis :

Kurtosis enables us to have an idea about the shape and nature the hump (middle part) of a frequency distribution. In other word Kurtosis is concerned with the flateness or peakness of the frequenc distribution.

As a measure of Kurtosis, Pearson introduced β_2 and γ_2 which ar

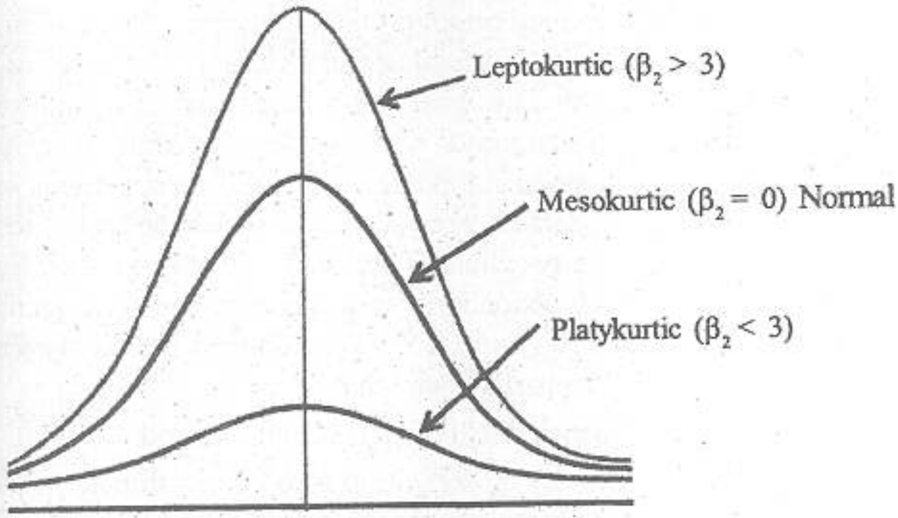
$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{\mu_4}{(\mu^2)^2}$$

$\gamma_2 = \beta_2 - 3$

If $\beta_2 = 3$ or $\gamma_2 = 0$ then it is mesokurtic (normal)

If $\beta_2 > 3$ or $\gamma_2 > 0$, then it is leptokurtic.

If $\beta_2 < 3$ or $\gamma_2 < 0$, then it is platykurtic.



Mean = Median = Mode

1.12 Computation of Moments for grouped frequency distribution:

In computing the moments about any arbitrary value as well as the central moments for a grouped frequency distribution, the calculations can be simplified by changing the scale.

Thus, for a grouped (continuous) frequency distribution rth moment about any arbitrary constant 'a' is given by,

$$\mu_r^a = \frac{1}{N}\Sigma f_i (x_i - a)^r$$

Now if the scale is changed, by introducing a new variable,

$$d_i = \frac{x_i - a}{h}$$

$$\mu_r^a = \frac{1}{N}\sum_i f_i (hd_i)^r = \frac{h^r}{N}\sum_i f_i d_i^r$$

Similarly, the scale may be changed while computing the central moments, we have

$$\mu_r^a = \frac{1}{N}\sum_i f_i (x_i - \mu)^r = \frac{1}{N}\sum_i f_i (hu_i)^r$$

$$= \frac{h^r}{N}\sum_i f_i u_i^r \left[\text{Putting } u_i = \frac{x_i - \mu}{h}\right]$$

1.13 Sheppard's Connection for Moments :

In case of grouped or continuous frequency distribution, for the calculation of moments, the value of the variable X is taken as the mid point of the corresponding class. This is based on the assumption that the frequencies are concentrated at the mid points of the corresponding classes. This assumption is approximately true for distributions which are symmetrical or moderately skewed and for which the class intervals are not greater than one-twentieth of the range of the distribution. However, in practice this assumption is not true in general and consequently some error known as 'grouping error' is introduced in the calculation of moments. W.F. Sheppard proved that if,

(1) The frequency distribution is continuous and

(2) The frequency tappers off to zero in both directions the effect due to grouping at the mid point of the intervals can be corrected by the following formula, known as Sheppard's correction,

$$\mu_2 \text{ (corrected)} = \mu_2 - \frac{h^2}{12}$$

$$\mu_3 \text{ (corrected)} = \mu_3$$

$$\mu_4 \text{ (corrected)} = \mu_4 - \frac{1}{2}h^2\mu_2 + \frac{7}{240}h^4$$

where, 'h' is the width of the class interval. The above corrections are valid only for symmetrical or slightly assymmetrical distributions. Moreover, as a safeguard against sampling errors, these should be perfectly applied only if the total frequency is fairly large, say greater than 1000.

1.14 Solved Examples :

Example 1. The first four moments of a distribution about the value 4 of the variable are –1.5, 17, –30 and 108. Find the moments about mean, β_1 and β_2. Find also the moments about (i) origin (ii) the point x = 2. [G.U. (MA/MSc) '00]

Solution :

Given, a = 4

$$\mu_1^a = -1.5, \mu_2^a = 17, \mu_3^a = -30, \mu_4^a = 108$$

The first four moments about mean are–

$$\mu_1 = 0$$

$$\mu_2 = \mu_2^a - \mu_1^{a^2} = 17 - (-1.5)^2 = 14.75$$

$$\mu_3 = \mu_3^a - 3\mu_2^a\mu_1^a + 2\mu_1^{a^3} = -30 - 3\times 17(-1.5) + 2(-1.5)^3 = 39.75$$

$$\mu_4 = \mu_4^a - 4\mu_3^a\mu_1^a + 6\mu_2^a\mu_1^{a^2} - 3\mu_1^{a^4}$$

$$= 108 - 4(-30)(-1.5) + 6 \times 17(-1.5)^2 - 3(-1.5)^4 = 142.31$$

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{(39.75)^2}{(14.75)^3} = 0.492$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{142.31}{(14.75)^2} = 0.654$$

(i) We have,

$$\mu_1^a = \mu - a$$

$$\Rightarrow -1.5 = \mu - 4 \quad \text{[since a = 4]}$$

$$\Rightarrow \mu = 2.5 \qquad \text{(which is the mean of the distribution)}$$

when a = 0 (i.e. origin),

$$\mu_1' = \mu = 2.5$$

$$\mu_2 = \mu_2' - \mu_1'^2 \Rightarrow 14.75 = \mu_2' - (2.5)^2 \Rightarrow \mu_2' = 21$$

Now, $\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3$

$$\Rightarrow 39.75 = \mu_3' - (3 \times 21 \times 2.5) + 2(2.5)^3 \Rightarrow \mu_3' = 166$$

Again $\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4$

$$\Rightarrow 142.31 = \mu_4' - (4 \times 166 \times 2.5) + (6 \times 21 \times 2.5^2) - 3(2.5)^4$$

$$\Rightarrow \mu_4' = 1131.99$$

(ii) We have, a = 2

$$\therefore \mu_1^a = \mu - a = 2.5 - 2 = 0.5$$

$$\mu_2 = \mu_2^a - \mu_1^{a^2} \Rightarrow 14.75 = \mu_2^a - (0.5)^2 \Rightarrow \mu_2^a = 15$$

$$\mu_3 = \mu_3^a - 3\mu_2^a\mu_1^a + 2\mu_1^{a^3} \Rightarrow 39.75 = \mu_3^a - (3 \times 15 \times 0.5) + 2(0.5)^3$$

$$\Rightarrow \mu_3^a = 62$$

$$\mu_4 = \mu_4^a - 4\mu_3^a\mu_1^a + 6\mu_2^a\mu_1^{a^2} - 3\mu_1^{a^4}$$

$$\Rightarrow 142.31 = \mu_4^a - (4 \times 62 \times 0.5) + (6 \times 15) \times (0.5)^2 - 3 \times (0.5)^4$$

$$\Rightarrow \mu_4^a = 243.99$$

Example 2. If the first, second and third raw moments are 2.5, 30 and 286 respectively, then what is the third central moment?

[G.U. (MA/MSc)Prev '00]

Solution : Given,

$$\mu_1' = 2.5, \quad \mu_2' = 30, \quad \mu_3' = 286$$

$$\therefore \mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2{\mu_1'}^3 = 286 - 3 \times 30 \times 2.5 + 2 \times (2.5)^3$$

$$= 92.25$$

Example 3. The first four moments of a distribution about a = 2 are 1, 2.5, 5.5 and 16. Calculate the first four moments about the origin.

[G.U. (MA/MSc) '03]

Solution :

Given, a = 2

$$\mu_1^a = 1, \quad \mu_2^a = 2.5, \quad \mu_3^a = 5.5, \quad \mu_4^a = 16.$$

We have, $\mu_1^a = \mu - a \Rightarrow 1 = \mu - 2 \Rightarrow \mu = 3$

$$\mu_2 = \mu_2^a - {\mu_1^a}^2 = 2.5 - 1^2 = 1.5$$

$$\mu_3 = \mu_3^a - 3\mu_2^a\mu_1^a + 2{\mu_1^a}^3 = 5.5 - (3 \times 2.5 \times 1) + 2 \times 1^3 = 0$$

$$\mu_4 = \mu_4^a - 4\mu_3^a\mu_1^a + 6\mu_2^a{\mu_1^a}^2 - 3{\mu_1^a}^4$$

$$= 16 - (4 \times 5.5 \times 1) + (6 \times 2.5 \times 1^2) - 3 \times 1^4 = 12$$

The first four moments about origin,

$$\mu_1' = \mu = 3$$

$$\mu_2 = \mu_2' - {\mu_1'}^2 \Rightarrow \mu_2' = 1.5 + 3^2 = 10.5$$

$$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2{\mu_1'}^3 \Rightarrow \mu_3' = 40.5$$

$$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'{\mu_1'}^2 - 3{\mu_1'}^4$$

$$\mu_4 = \mu_4' - 4 \times 40.5 \times 3 + 6 \times 10.5 \times 3^2 - 3 \times 3^4$$

$$\Rightarrow \mu_4' = 174$$

Example 4. On the basis of the following data calculate the first four moments about the value 15, about mean and also β_1 and β_2

CI :	0–10	10–20	20–30	30–40
f :	1	3	4	2

CI	f	Mid value	$d = \frac{x-a}{h}$	fd	fd^2	fd^3	fd^4
0–10	1	5	−1	−1	1	−1	1
10–20	3	15	0	0	0	0	0
20–30	4	25	1	4	4	4	4
30–40	2	35	2	4	8	16	32
	N=10			$\sum fd = 7$	$\sum fd^2 = 13$	$\sum fd^3 = 19$	$\sum fd^4 = 37$

We have $\mu_r^{a} = h^r \frac{1}{N} \Sigma fd^r$ where h = 10, a = 15

$$\therefore \mu_1^{a} = h \frac{1}{N} \Sigma fd = 10 \frac{1}{10} \times 7 = 7$$

$$\mu_2^{a} = h^2 \frac{1}{N} \Sigma fd^2 = 10^2 \frac{1}{10} \times 13 = 130$$

$$\mu_3^{a} = h^3 \frac{1}{N} \Sigma fd^3 = 10^3 \frac{1}{10} \times 19 = 1900$$

$$\mu_4^{a} = h^4 \frac{1}{N} \Sigma fd^4 = 10^4 \frac{1}{10} \times 37 = 37000$$

$$\mu_2 = \mu_2^{a} - \mu_1^{a^2} = 130 - 7^2 = 81$$

$$\mu_3 = \mu_3^{a} - 3\mu_2^{a}\mu_1^{a} + 2\mu_1^{a^3} = 1900 - 3 \times 130 \times 7 + 2 \times 7^3 = -144$$

$$\mu_4 = \mu_4^{a} - 3\mu_3^{a}\mu_1^{a} + 6\mu_2^{a}\mu_1^{a^2} - 3\mu_1^{a^4}$$

$$= 37000 - 4 \times 1900 \times 7 + 6 \times 130 \times 7^2 - 3 \times 7^4 = 14817$$

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{(-144)^2}{(81)^3} = 0.039$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{37000}{130^2} = 2.2$$

$\beta_1 = 0.039$. Hence given distribution is slightly assymetric distribution.

$\beta_2 = 2.2 < 3$, Hence the given distribution is platykurtic.

Example 5. The first four moments of a series about the value 5 are 1, 5, 10 and 50 respectively. Obtain the various characteristics of the distribution on the basis of the information given. Comment upon nature of the distribution. [GU(MA/MSc) '0

Solution : Given,

$$a = 5,\ \mu_1^a = 1, \mu_2^a = 5, \mu_3^a = 10, \mu_4^a = 50$$

We have, $\mu_1^a = \mu - a \Rightarrow 1 = \mu - 5 \Rightarrow \mu = 6$

$$\mu_2 = \mu_2^a - \mu_1^{a^2} = 5 - 1^2 = 4$$

$$\mu_3 = \mu_3^a - 3\mu_2^a\mu_1^a + 2\mu_1^{a^3} = 10 - 3\times5\times1 + 2\times1 = -3$$

$$\mu_4 = \mu_4^a - 4\mu_3^a\mu_1^a + 6\mu_2^a\mu_1^{a^2} - 3\mu_1^{a^4}$$

$$= 50 - 4\times10\times1 + 6\times5\times1^2 - 3\times1^4 = 37$$

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{(-3)^2}{4^3} = 0.14$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{37}{4^2} = 2.31$$

Again,

$$\gamma_1 = \frac{\mu_3}{\sigma^3} = \frac{-3}{2^3} < 0 \qquad \left[\because\ \sigma^2 = \mu_2 = 4 \Rightarrow \sigma = 2\right]$$

Since $\gamma_1 < 0$, the distribution is negatively skewed. Again $\beta_2 <$ Hence the given distribution is platykurtic.

Example 6. Arithmetic mean of a certain distribution is 5. The seco and third moments about the mean are 20 and 140 respectively. Find t second and third moments about 10. [G.U.(MA/MSc) '0

Solution :

Given, $\mu = 5$, $\mu_2 = 20$, $\mu_3 = 140$

We have,

$$\mu_1^a = \mu - a = 5 - 10 = -5$$

$$\mu_2 = \mu_2^a - \mu_1^{a^2} \Rightarrow 20 = \mu_2^a - (-5)^2 \Rightarrow \mu_2^a = 45$$

$$\mu_3 = \mu_3^a - 3\mu_2^a\mu_1^a + 2\mu_1^{a^3} \Rightarrow 140 = \mu_3^a - 3\times45\times(-5) + 2(-5)^3$$

$$\Rightarrow \mu_3^a = -285$$

CHECK YOUR PROGRESS

1. In the theory of central moment show that

$$\mu_r = \mu_r' - {}^rC_1\mu_{r-1}'\mu_1' + {}^rC_2\mu_{r-2}'\mu_1'^2 ......+(-1)^2\mu_1'^r$$

[GU(MA/MSc) Prev '99]

2. The first three moments of a distribution about the value 2 are 1, 16, –40. Find mean and variance.

[GU(MA/MSc) Prev '99]

3. Express the moments about any arbitrary constant 'a' in terms of moments about the mean. [GU(MA/MSc) Prev '02]

4. Show that the central moments are invariant to the change of origin but not of scale.

   Calculate the value of β_1 and β_2 from the following data and comment on the result.

Marks :	20–30	30–40	40–50	50–60	60–70	70–80	80–90
No. of students :	4	7	10	20	4	3	2

[GU(MA/MSc) Prev '01]

5. Define moments. Derive the relationship between moments about the mean and the moments about the origin. Hence express the first four central moments in terms of the moments about the origin.

1.15 Summing Up :

Probability is the measure of likelihood of occurrances of chance events. Probability gives a quantitative measure of the chance of a random experiment. There are 3 approaches to probability viz. classical approach, empirical approach and axiomatic approach.

Conditional probability is the probability attached to events which are such that one of them occurs only when the other is known to have occurred.

Events are said to be independent of each other if happening of any one of them is not affected by and does not affect the happening of any one of the other.

A random variable is a numerical valued function defined on a sample space. The probability function of a discrete random variable is known as the 'probability mass function' and that of a continuous random variable is known as the 'probability density functions'.

Mathematical expectation is the summation of the products of the various values that a random variable may take and their respective probabilities.

Moment describes the various characteristics of a frequen distributions like central tendency, variation, skewness and kurtosis et

Skewness refers to lack of symmetry i.e., when a distribution is n symmetrical, it is called a skewed distribution. Kurtosis enables us have an idea about the shape and nature of the hump (middle part) of frequency distribution.

1.16 References and Suggested Readings :

1. Gupta, S.C. and Kapoor, U.K., "Fundamentals of Mathematic Statistics."
2. Gupta, S.C. and Kapoor, V.K. "Fundamentals of Applie Statistics."
3. Gupta, S.C., "Fundamentals of Statistics."
4. Agarwal, D.R., "Business Statistics."
5. Hazarika, P, "Essential Statistics for Economics and Commerce."

Unit 2 : Standard Probability Distribution

Contents :

2.1 Introduction :

The distributions which are derived mathematically based on certain assumptions are called Theoretical Distributions Or Expected Destributions or Probability Distributions. A probability distribution whose underlying variable is discrete is called a discrete probability distribution; on the other hand, a probability distribution is called a continuous probability distribution if its underlying random variable is a continuous one. Here we shall discuss three important probability distribution viz Binomial distribution, Poisson distribution and Normal distribution. The first two are discrete probability distribution and the third is a continuous probability distribution. In the Unit-1 you have already read about the discrete and continuous random variables. In this unit you will learn some more about their pattern of distribution.

2.2 Objectives :

This unit is designed to help you understand the standard probabilit distributions. After reading this unit you will be able to,

1. Distinguish between the pattern of probability in case of small an large samples.

2. Depending on the normal distribution predict the characterstic o a sample for the total relevant population.

3. Calculate various probabilistic situations through binomial poisson and normal distributions.

4. Define moment generating function and the central limit theory.

2.3 Binomial Distribution (or Bernoulli Distribution) :

Binomial distribution, which is also known as the 'Bernoull distribution' is the most fundamental and important discrete probabilit distribution in statistics and is defined by the probability function–

$P(X) = {}^nC_x p^x q^{n-x};$ $\quad X = 0, 1, 2, \ldots, n$ —(i)

which gives the probability of X successes in a series of n independent trials ($x \leq n$) where p is the constant probability of success in a single trial and $q = 1 - p$.

2.4 Assumptions of Binomial Distribution :

Binomial distribution is based on the following assumptions :

(i) An experiment is performed under identical conditions for a finite and fixed number of trials (n).

(ii) Each trial must result in two mutually exclusive outcomes - success or failure.

(iii) In binomial distribution, the outcome of any trial does not affect the outcomes of the subsequent trials. Hence trials are independent.

(iv) In each trial, the probability of success p remains constant.

For given n and p, the above function p(x) represents the following discrete probability distribution.

X	0	1	2	....	n
p(X)	q^n	${}^nC_1 p^1 q^{n-1}$	${}^nC_2 p^2 q^{n-2}$	....	p^n

The probability of 0 success, 1 success, 2 successes, ..., n successes are respectively the 1st, 2nd, 3rd, ... , (n + 1) th terms of the Binomial expansion. $(q + p)^n = q^n + {}^nC_1 p^1 q^{n-1} + {}^nC_2 p^2 q^{n-2} + \ldots + p^n$.

This is why the probability function (i) is called a Binomial Probability Distribution.

2.5 Properties of Binomial Distribution :

(1) n and p are called the parameters of the Binomial distribution. The binomial distribution is completely known when the values of n and p are known.

(2) Binomial distribution is a discrete probability distribution in which the random variable X (the number of success) assumes the values 0, 1, 2, ..., n where n is finite.

(3) The sum of all the probabilities of successess is unity i.e.

$$P(0) + P(1) + P(2) + ... + P(n) = 1$$

(4) The mean or expected value of the binomial distribution is np,

Standard deviation is $\sqrt{npq}$,

Variance is npq

$$\beta_1 = \frac{(q-p)^2}{npq}$$

$$\beta_2 = 3 + \frac{1-6pq}{npq}$$

(5) The mode of binomial distribution is that value of the random variable X which occurs with the largest probability. It may have either one or two modes.

(6) If two independent random variables X and Y follow binomial distribution with parameters (n_1, p) and (n_2, p) respectively, their sum $(X+Y)$ also follows binomial distribution with parameters $(n_1 + n_2, p)$.

(7) The binomial distribution is a theoretical distribution rather than, an observed frequency distribution.

(8) The shape of a binomial distribution depends on the values of p, q and n.

(9) The binomial distribution can be presented graphically by means of a line graph. The number of successes (x) is taken on X axis and the probability of successes on the Y axis.

STOP TO CONSIDER

(1) If n independent trials are repeated N times, we get N sets of n trials and the expected frequency of X successes is $N.{}^nC_x p^x q^{n-x}$.

The expected frequencies of 0, 1, 2, ..., n successes are the successive terms of binomial expansion of $N(q + p)^n$.

(2) The parameter(s) of a distribution (is) are the quantity (quantities) which, when known, the distribution is completely known.

2.6 Solved Examples :

Example 1. Show that in Binomial distribution Mean > Variance.

Solution :

Let, $X \sim B(n, p)$

We have, $p + q = 1$

$\therefore q < 1$

$\Rightarrow npq < np$ (multiplying both sides by np)

$\Rightarrow$ Variance < Mean

$\Rightarrow$ Mean > Variance.

Example 2. Bring out fallacy, if any, in the statement, "The mean of the Binomial distribution is 5 and standard deviation is 3."

[G.U.(MA/MSc)Prev. '96]

Solution :

According to the statment,

$$np = 5 \text{ ....(i)}$$

$$\sqrt{npq} = 3 \text{ ....(ii)}$$

$$\Rightarrow npq = 9 \text{ .... (iii)}$$

Now,

$$\text{(i)} \div \text{(iii)} \Rightarrow \frac{1}{q} = \frac{5}{9}$$

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

Here $q > 1$

$\therefore$ The given statement is incorrect.

Example 3. X is a random variable following the probability law $P(X = X) = {}^nC_x P^x(1-p)^{n-x}$ where $0 \le p \le 1$ and $x = 0, 1, ..., n$. If mean of X is 5, variance of X is 4. Find the value of n. [G.U.(MA/MSc)'99]

Solution :

Let, $X \sim B(n, p)$

Given, Mean = np = 5, Variance = npq = 4

$$\therefore \frac{npq}{np} = \frac{4}{5}$$

$$\Rightarrow q = \frac{4}{5} \quad \Rightarrow p = 1 - \frac{4}{5} = \frac{1}{5}$$

Again, np = 5

$$\Rightarrow n \tfrac{1}{5} = 5 \Rightarrow n = 5 \times 5 = 25$$

Example 4. Suppose, X is a random variable following the Binomial probability law. If the mean of X is 5, the variance of X cannot be larger than 5. Explain why it is so? [G.U.(MA/MSc.)'00]

Solution :

Let, $X \sim B(n, p)$

Given, Mean = np = 5

Variance = npq = 6 (say i.e. larger than 5)

$$\therefore \frac{npq}{np} = \frac{6}{5}$$

$$\Rightarrow q = 1.2 > 1$$

which is impossible. Hence the variance cannot be larger than 5.

Example 5. The mean of binomial distribution is 20 and standard deviation is 4, calculate n, p and q. [G.U.(MA/MSc)'02]

Solution :

Given, Mean (np) = 20

Variance (npq) = 4^2 = 16

$$\frac{npq}{np} = \frac{16}{20} \Rightarrow q = 0.8 \Rightarrow p = 1 - 0.8 = 0.2$$

Again, np = 20 $\Rightarrow$ n(0.2) = 20 $\Rightarrow$ n = 100

Example 6. In a biochemical experiment 20 insects were put into each of 100 jars and were subjected to fumigant. After 3 hours the number of living insects in each jar was counted and the results were as follows :

No. of insects alive :	0	1	2	3	4	5	6	7	8	9
No. of jars :	3	8	11	15	16	14	12	11	9	1

Fit a binomial distribution to the above data.[G.U.(MA/MSc)'04]

Solution :

Let, X denotes the number of insects alive in a jar.

and again, Let, $X \sim B(n, p)$

Given,

$n = 20, N = 100$

X	Observed Frequency (f)	fx	P(x)	Expecte frequen
0	3	0	${}^{20}C_0(.22)^0\ (.78)^{20} = .006$	0.6 =
1	8	8	${}^{20}C_1(.22)^1\ (.78)^{19} = .039$	3.9 =
2	11	22	${}^{20}C_2(.22)^2\ (.78)^{18} = .105$	10.5 =
3	15	45	${}^{20}C_3(.22)^3\ (.78)^{17} = .178$	17.8 =
4	16	64	${}^{20}C_4(.22)^4\ (.78)^{16} = .213$	21.3 =
5	14	70	${}^{20}C_5(.22)^5\ (.78)^{15} = .192$	19.2 =
6	12	72	${}^{20}C_6(.22)^6\ (.78)^{14} = .136$	13.6 =
7	11	77	${}^{20}C_7(.22)^7\ (.78)^{13} = .076$	7.6 =
8	9	72	${}^{20}C_8(.22)^8\ (.78)^{12} = .035$	3.5 =
9	1	9	${}^{20}C_9(.22)^9\ (.78)^{11} = .013$	1.3 =
	N = 100	$\sum fx = 439$		

$$\bar{x} = \frac{\Sigma fx}{N} = \frac{439}{100} = 4.39$$

In binomial distribution Mean = np

$\Rightarrow 4.39 = 20p \Rightarrow p = .22 \Rightarrow q = .78$

$\therefore P(X = r) = {}^{n}C_r p^r q^{n-r}$

$= {}^{20}C_r(.22)^r\ (.78)^{20-r}$

Example 7. A multiple choice test consists of 8 questions with 3 answ to each questions of which only one is correct. A student answers ea question by rolling a balanced die. He gives a tick mark to the fi answer if he gets 1 or 2, the second answer if he gets 3 or 4 and third answer if he gets 5 or 6. To get a distinction the student must sec at least 75% correct answer. If there is no negative marking what is probability that the student secure a distinction. [G.U. (MA/MSc)'(

Solution :

Let X denotes the number of correct answers.

Let, $X \sim B(n, p)$

Given,

$$n = 8, \qquad p = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}, \qquad q = 1 - \frac{1}{3} = \frac{2}{3}$$

Now, 75% correct answers out of 8 questions is,

$$8 \times 0.75 = 6.$$

$$\therefore P(X = r) = {}^{n}C_{r}\, p^{r} q^{n-r}$$

$$= {}^{8}C_{r}\left(\frac{1}{3}\right)^{r}\left(\frac{2}{3}\right)^{n-r}, \qquad r = 0, 1, \ldots.. 8$$

$$P(X \geq 6) = P(X = 6) + P(X = 7) + P(X = 8)$$

$$= {}^{8}C_{6}\left(\tfrac{1}{3}\right)^{6}\left(\tfrac{2}{3}\right)^{2} + {}^{8}C_{7}\left(\tfrac{1}{3}\right)^{7}\left(\tfrac{2}{3}\right)^{1} + {}^{8}C_{8}\left(\tfrac{1}{3}\right)^{8}\left(\tfrac{2}{3}\right)^{0}$$

$$= 0.0187$$

$\therefore$ The student has 0.0187 probability to secure a distinction.

CHECK YOUR PROGRESS

1. What is a binomial distribution? What are its mean and standard deviation? A coin is tossed 4 times. What is the probability of obtaining 2 or more heads? [G.U.(MA/MSc)Prev '06]
2. What is binomial distribution? What are its properties? [G.U.(MA/MSc)Prev '07]
3. A student obtained the following result. Is the result consistent? For a binomial distribution mean = 4 variance = 6. [G.U.(MA/MSc) '97]
4. For a binomial distribution the mean is 4 and variance is 2. Find the probability of getting (i) at least 2 successes (ii) at most 2 successes.
5. In a binomial distribution consisting of S independent trials the probability of 1 and 2 successes are 0.4096 and 0.2048 respectively. Find the parameter P of the distribution and also obtain mode and variance.
6. 6 dice are thrown 729 times. How many times of you expect at least 3 dice shows a 5 or 6?
7. If on an average one ship in every 10 sinks, find the probability that out of 5 ships, at least 4 will arive safely.
8. The mean of a binomial distribution is 20 and standard deviation is 4. Find out the characteristics of the distribution.
9. For a Binomial variate X the mean = 4 and vaniance = 3, find P(X = a non zero value).

Let us understand, Binomial Theorem :

P(n, k, p) means

- Probability of k successes in n trials where the probability of success on any one trial is p.
- k specified outcomes.
- n trials
- p probability of the specified outcome in one trial.

2.7 Poisson Distribution :

Simon D Poisson (1781-1840) had developed the Poiss distribution as a limiting case of Binomial probability distribution under t following conditions.

(i) n, the number of trials is indefinitely large i.e. $n \to \infty$

(ii) p, the constant probability of success for each trial is indefinite small i.e. $p \to 0$

(iii) np = m (say) is finite.

Under the above three conditions the Binomial probability functio tends to the probability function of the Poisson distribution given belo

$$p(r) = P(X = r) = \frac{e^{-m} . m^r}{r!}, \; r = 0, 1, 2, 3, ...$$

Poisson distribution can be used to explain the behaviour of the discre random variables such that the probability of occurance of the event very small and the total number of possible cases is sufficiently large.

2.8 Properties of Poisson Distribution :

1. Poisson distribution is a discrete probability distribution since t variable X can take only integral values i.e. 0, 1, 2, ... α.

2. The sum of all probabilities of successes is unity i.e. tot probability is one.

$$\sum_{r=0}^{\alpha} P(r) = e^{-m} + me^{-m} + \frac{m^2}{2!}e^{-m} + \frac{m^3}{3!}e^{-m} + ...$$

$$= e^{-m}\left[1 + m + \frac{m^2}{2!} + \frac{m^3}{3!} + ...\right]$$

$$= e^{-m} \times e^{m} \qquad \left[\because e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + ...\right]$$

$$= e^{-m+m} = e^0 = 1 \text{ [by law of indices)}$$

Poisson Probabilities

No. of Successes r	Probability P(r)
0	$\frac{e^{-m}.m^0}{0!} = e^{-m}$
1	$\frac{e^{-m}.m}{1!}$
2	$\frac{e^{-m}.m^2}{2!}$
3	$\frac{e^{-m}.m^3}{3!}$

(3) 'm' is the only parameter of the Poisson distribution. Hence, if we know the value of m, all the probabilities of the Poisson distribution can be obtained.

(4) The mean or expected value of poisson distribution is 'm', standard deviation $= \sqrt{m}$.

$\therefore$ Variance = m

$$\beta_1 = \frac{1}{m}$$

$$\beta_2 = 3 + \frac{1}{m}$$

5. If X and Y are two independent poisson variable with parameters m_1 and m_2 respectively, then their sum X + Y is also a poisson with parameters $m_1 + m_2$.

6. It is limiting form of the binomial distribution.

7. Mode in the poisson distribution can be obtained under the following two cases.

Case I : When m is not an integer

Let, m = k + f where k is the integer and f is the fractional value i.e. $0 < f < 1$.

In this case, the mode of the poisson distribution is k.

Case II : When m is an integer.

Let, m = k where k is the integer value.

In this case you will get 2 values of mode i.e. k and k – 1.

STOP TO CONSIDER

Role/Uses/Importance of Poisson Distributions :

1. It is used in quality control statistics to count the number of defects of an item.
2. In biology to count the number of bacteria.
3. In physics to count the number of particles emitted from a radio-active substance.
4. In insurance problems to count the number of causalities.
5. In a waiting time problems to count the number of incoming phone calls or incoming customers.
6. Number of traffic arrivals such as trucks at terminals, aeroplanes at airports, ships at docks, and so forth.
7. In determining the number of death in a district in a given period say, a year, by a rare disease.
8. The number of typographical errors per page in typed material, number of deaths as a result of road accidents.
9. In problems dealing with the inspection of manufactured products with the probability that any one piece is defective, is very small and the lots are very large.
10. To model the distribution of the number of persons joining a que (a line) to receive a service or purchase of a product etc.

2.9 Solved Examples :

Example 1. A manufacturer of copper pins knows that 2% of his product is defective. If he sells copper pins in boxes of 200 and guarantees that not more than 5 pins will be defective. What is the probability that a box will fail to meet the guranteed quality?

(Given $e^{-4} = 0.0183$) [G.U.(MA/MSc)Prev'98]

Solution :

Let X denotes the number of defective copper pins in a box.

We assume that, X follows a poisson distribution with mean, m = np = 200 × 0.02 = 4

Then,

$$P(X) = \frac{e^{-m}.m^{r}}{r!}; \qquad r = 0, 1, 2, ...$$

$$= \frac{e^{-4}.4^{r}}{4!}$$

The box will fail to meet the guaranteed quality if the number of defective items in the box exceeds 5.

Now,

$$P(X \geq 5) = 1 - [P(0) + P(1) + P(2) + P(3) + P(4) + P(5)]$$

$$= 1 - 0.0183\left[1+4+8+\frac{32}{3}+\frac{32}{3}+\frac{128}{15}\right]$$

$$= 0.2155.$$

Example 2. What probability model is appropriate to describe a situation where 100 misprints are distributed randomly throught the 100 pages of a book ? For this model what is the probability that a page is taken at random will contain,

(i) At most two misprints.

(ii) At least 3 misprints. [GU(MA/MSc)'00]

Solution :

The given problem can be suitably described by the Poisson distribution.

Let, X denotes the number of misprint in a page.

Let, $X \sim P(m)$

Given,

number of misprints in 100 pages = 100

$$\text{Number of misprints in 1 page} = \frac{100}{100} = 1$$

Average number of misprint = m = 1

$$P(X = r = \frac{e^{-m}.m^{r}}{r!}; r = 0, 1, 2, ...$$

(i) The probability that at most two misprints are there,

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$$

$$P(X = Y) = \frac{e^{-1}.1^{0}}{0!}+\frac{e^{-1}.1^{1}}{1!}+\frac{e^{-1}.1^{2}}{2!}$$

$$= e^{-1}\left(1+1+\tfrac{1}{2}\right) = 0.92$$

(ii) The required probability that at least 3 misprints are there is,

$$P(X \geq 3) = P(X = 3) + P(X = 4) + P(X = 5) + ...$$

$$= 1 - [P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)]$$

$$= 1 - \frac{e^{-1}.1^{0}}{0!} - \frac{e^{-1}.1^{1}}{1!} - \frac{e^{-1}.1^{2}}{2!} - \frac{e^{-1}.1^{3}}{3!}$$

$$= 1 - \{0.368 \times 2.667\} = 0.02$$

Example 3. A car hire firm has 2 cars which it hires out day by da[y]
The number of demand for a car on each day is distributed as a poisso[n]
variate with mean 1.5. Calculate the proportion of days on which.

(i) Neither car is used.

(ii) Some demand is refused. [G.U. (M.A./MSc)'00]

Solution :

Let, X denotes the number of demand for the two cars in a day

Let, $X \sim P(m)$

Given, $m = 1.5$

$$\therefore P(X = r) = \frac{e^{-m}.m^r}{r!}, r = 0, 1, 2, ...$$

$$= \frac{e^{-1.5}.1.5^r}{r!}$$

(i) Required probability,

$$P(X = 0) = \frac{e^{-1.5}.1.5^0}{0!} = 0.2231$$

(ii) Required probability,

$$\begin{aligned} P(X > 2) &= P(X = 3) + P(X = 4) + P(X = 5) + ... \\ &= 1 - P(X = 0) - P(X = 1) - P(X = 2) \\ &= 1 - \frac{e^{-1.5}.1.5^0}{0!} - \frac{e^{-1.5}.1.5^1}{1!} - \frac{e^{-1.5}.1.5^2}{2!} \\ &= 1 - e^{-1.5}(1 + 1.5 + 1.125) = 0.192 \end{aligned}$$

Example 4. If X is a poisson variate such that

$P(X = 2) = 9\ P(X = 4) + 90\ P(X = 6)$

Find (i) Mean and standard deviation

(ii) Coeffecient of skewness and Kurtosis. [G.U. (MA/MSc)'01]

Solution : Let, $X \sim P(m)$

$$P(X = 2) = 9\ P(X = 4) + 90\ P(X = 6)$$

$$\Rightarrow \frac{e^{-m}.m^2}{2!} = 9\frac{e^{-m}.m^4}{4!} + 90\frac{e^{-m}.m^6}{6!}$$

$$\Rightarrow 1 = \frac{3m^2}{4} + \frac{m^4}{4} \quad \left[\text{dividing both sides by } \frac{e^{-m}.m^2}{2\,!}\right]$$

$$\Rightarrow 4 = 3m^2 + m^4$$

$$\Rightarrow m^4 + 3m^2 - 4 = 0 \Rightarrow m^4 + 4m^2 - m^2 - 4 = 0$$

$$\Rightarrow m^2\left(m^2+4\right) - 1\left(m^2+4\right) = 0$$

$$\Rightarrow \left(m^2+4\right)\left(m^2-1\right) = 0$$

$\Rightarrow m^2 + 4 = 0$	Or $m^2 - 1 = 0$
$\Rightarrow m^2 = -4$	$\Rightarrow m^2 = 1$
Which is impossible	$\Rightarrow m = \pm 1$

$$\therefore m = 1 \; (\because m > 0)$$

(i) Mean = m = 1, Variance = m = 1

Standard Deviation $= \sqrt{m} = 1$

(ii) Coeffecient of Skewness $(\beta_1) = \dfrac{1}{m} = \dfrac{1}{1} = 1$

Coeffecient of Kurtosis $(\beta_2) = 3 + \dfrac{1}{m} = 3 + 1 = 4$

Example 5. If X and Y are independent poisson distribution such that

$P(X = 1) = P(X = 2)$ and $P(Y = 2) = P(Y = 3)$.

Find Var (X – 2Y).

[G.U. (MA/MSc)'04]

Solution : Let, $X \sim P(m)$ and $Y \sim P(\lambda)$

$$\therefore P(X = r) = \frac{e^{-m}.m^r}{r!} \text{ and } P(Y = r) = \frac{e^{-\lambda}.\lambda^r}{r!}; r = 0, 1, 2, \ldots$$

Given, $P(X = 1) = P(X = 2)$;	$P(Y = 2) = P(Y = 3)$
$\Rightarrow \dfrac{e^{-m}.m}{1!} = \dfrac{e^{-m}.m^2}{2!}$	$\Rightarrow \dfrac{e^{-\lambda}.\lambda^2}{2!} = \dfrac{e^{-\lambda}.\lambda^3}{3!}$
$\Rightarrow 1 = \dfrac{m}{2}$	$\Rightarrow 1 = \dfrac{\lambda}{3}$
$\Rightarrow m = 2$	$\Rightarrow \lambda = 3$

Var (X = 2) and Var (Y = 3)

$$\text{Var}(X - 2Y) = E[(X - 2Y) - E(X - 2Y)]^2$$
$$= E[X - 2Y - E(X) + 2E(Y)]^2$$
$$= E[\{X - E(X)\} - 2\{Y - E(Y)\}]^2$$
$$= E[\{X - E(X)\}^2 + 4\{Y - E(Y)\}^2 - 4\{X - E(X)\}\{Y - E(Y)\}]$$
$$= E[X - E(X)]^2 + 4\,E[Y - E(Y)]^2 - 4\,E[\{X - E(X)\}\{Y - E(Y)\}]$$
$$= \text{Var}(X) + 4\,\text{Var}(Y) - 4\,\text{Cov}(X, Y)$$

$= 2 + (4 \times 3) - (4 \times 0)$

($\because$ Cov (X, Y) = 0 When X and Y are independent

$= 14$

Example 6. If X is a poisson variate and $P(X = 0) = P(X = 1) = k$. Find the value of k. [GU(MA/MSC)'5]

Solution : Let, $X \sim P(m)$

$$P(X = 0) = \frac{e^{-m}.m^r}{r!} \ ; r = 0, 1, 2, ...$$

Given, $P(X = 0) = P(X = 1) = k$

$$\Rightarrow \frac{e^{-m}.m^0}{0!} = \frac{e^{-m}.m}{1!} = k \Rightarrow 1 = m = k \Rightarrow k = 1$$

Let us understand,

The equality of mean and variance is an important characterstic of the Poisson distribution, whereas for the binomial distribution the mean is always greater than the variance.

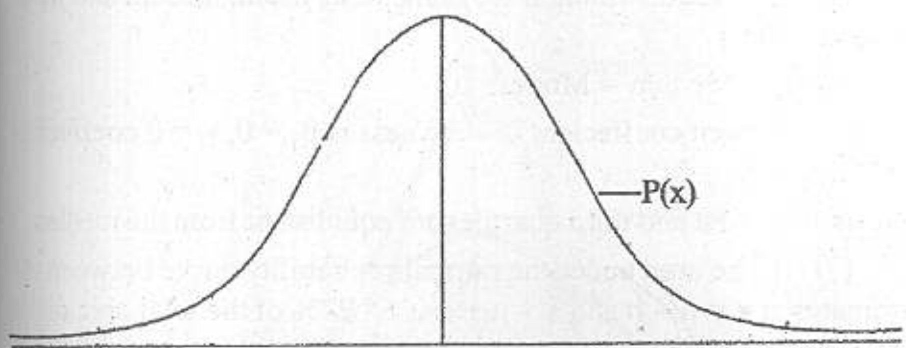
CHECK YOUR PROGRESS

1. Distinguish between Binomial and Poisson probability distributions. Give their characteristics with examples. [G.U.(MA/MSc)Prev '09]
2. Define Poisson distribution. What are its properties? [G.U.(MA/MSc)Prev '09]
3. It has been observed that average number of suicides per week in Delhi in 1.5. If X denotes the number of suicide cases in a month, find $P(X \geq 5)$. [G.U.(MA/MSc)Prev '96]
4. In a Poisson distribution the probability P(x) for x = 0 is 10% find
   (i) The mean of the distribution [G.U.(MA/MSc)Prev '97]
   (ii) The mode of the distribution
   (iii) The coeffecient of skewness and Kurtosis.
5. If the probability that an individual suffer a bad reaction from a particular injection is 0.001. Determine the probability that out of 2000 individuals,
   (a) Exactly 3, (ii) More than 2 (iii) Atleast 3 individuals will suffer from a bad reaction.
6. The variance of a poisson distribution is 4. Find the probability that X = 3 (Given $e^{-4} = 0.0183$).
7. If 2% of electric bulbs manufactured by a certain company are defective, find the probability that in sample of 200 bulbs (i) less than 2 bulbs (ii) more than 3 bulbs are defective (Given e^{-4} = 0.0183).

2.10 Normal Distribution :

Normal distribution is a continuous probability distribution. A continuous distribution is one in which the underlying variable X (say) may assume any value within a given range. A continuous variable is usually represented by a smooth, bell shaped curve which is perfectly symmetric. There are many continuous distributions of which the most commonly used distribution is normal distribution.

The normal distribution was first discovered by De-Moivre (1667-1754) in 1733. Normal distribution is also known as Gaussian distribution (Gaussian Law of Errors) after Karl Friedrich Gauss (1777-1855) who used his distribution to describe the theory of accidental errors of measurements involved in the calculation of orbits of heavenly bodies.



Mean = μ

Normal Curve

If x is a continuous random variable following Normal probability distribution with mean μ and standard deviation σ, then its probability density function (p.d.f) is given by,

$$P(X) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2}, \quad -\infty < X < \infty \quad ....(i)$$

$$\Rightarrow P(X) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(X-\mu)^2}{2\sigma^2}}, \quad -\infty < X < \infty$$

where, $\pi = \frac{22}{7}, \sqrt{2\pi} = 2.5066$

$e = 2.71828$

Note : $\int_a^b f(x)dx$ gives the probability that the value of X lies between

a and b. This is equal to the area under the normal curve between t
two ordinates at X = a and X = b on the x axis. The total area und
the normal curve is unity

$$P(a \leq X \leq b) = \int_{a}^{b} f(x)dx \quad \text{and} \quad \int_{-\alpha}^{\alpha} f(x)dx = 1$$

2.11. Properties of Normal Distribution :

(1) Normal distribution is a continuous probability distribution havi two parameters i.e. mean μ and standard deviation σ.

(2) The graph of p(x) is the famous bell shaped curve with top o the bell directly above mean (μ).

(3) The distribution is symmetrical about the line $x = \mu$ and two tai of the curve on both sides of $x = \mu$ extends to infinity.

(4) Since the distribution is symmetrical, mean, median and mod coincide. Thus,

Mean = Median = Mode = μ.

(5) Moment coeffecient of skewness is $\beta_1 = 0, \gamma_1 = 0$ coeffecien of Kurtosis $\beta_2 = 0$.

(6) The first and third quartiles are equidistant from the median.

(7) (i) The area under the normal probability curve between the ordinates at $x = \mu - \sigma$ and $x = \mu + \sigma$ is 68.27% of the total area unde the curve. Since the total area is unity, therefore, the area = 0.6827.

(ii) The area under the normal probability curve between the ordinates at $x = \mu - 2\sigma$ and $x = \mu + 2\sigma$ is 95.45%.

(iii) The area under the normal probability curve between the ordinates at $x = \mu - 3\sigma$ and $x = \mu + 3\sigma$ is 99.73% of the total area under the curve. This area is 0.9973.

2.12 Importance of Normal Distribution :

Normal probability distribution or commonly called the normal distribution is one of the most important continuous theoretical distribution of statistics. Most of the data relating to Economics and Business Statistics or even in social and physical sciences conform to this distribution. It has been found that :

(1) Data obtained from Psychological, Physical and Biological measurements approximately follow Normal distributions.

(2) Distributions like Binomial, Poisson etc can be approximated to Normal distribution.

(3) Normal curve is used to find confidence limits of the population parameters.

(4) Normal distribution is largely applied in statistical quality control

in industry for finding control limits.

(5) The theory of errors of observations in physical measurements are based on normal distribution.

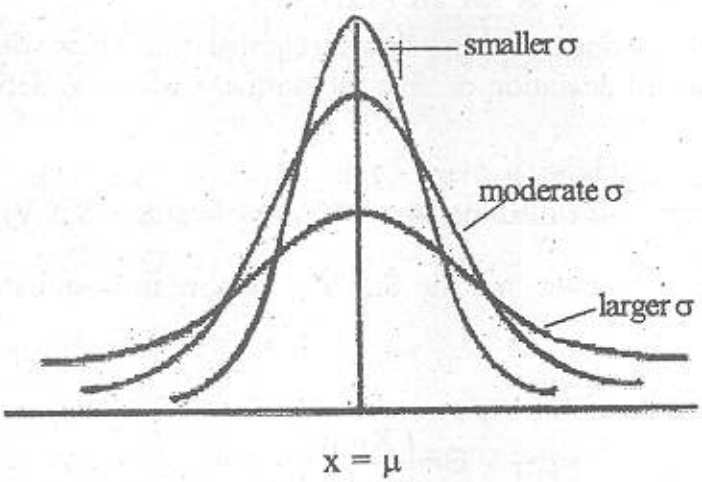
STOP TO CONSIDER

1. The sum as well as difference of independent normal variates is also normal variate.

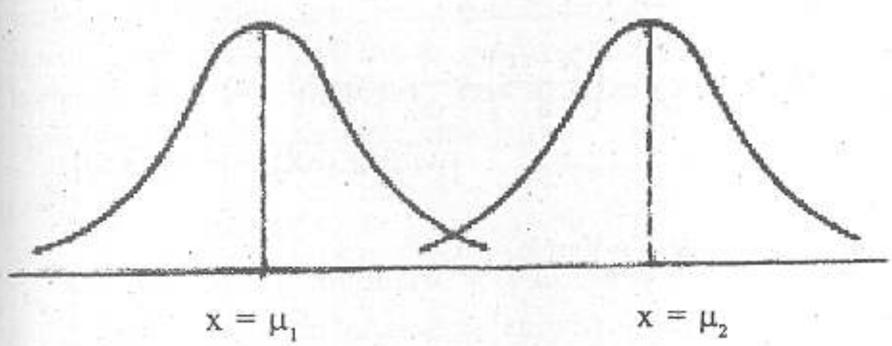
2. The means and standard deviations of two or more normal distributions are independent of each other. For two normal curves, it is not necessary that a normal curve with larger (smaller) mean will also have larger (smaller) standard deviaton.

You may get,

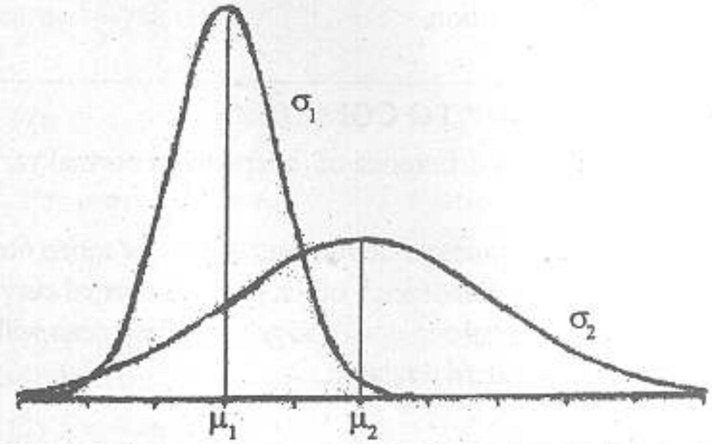
(i) The same means and different standard deviations.



(ii) The same standard deviation but different means.



(iii) Different means and different standard deviations.



2.13 Standard Normal Variate (Variable) :

If X is a random variable following Normal distribution with mean μ and standard deviation σ, then the normal variable Z defined as follows,

$Z = \frac{X-\mu}{\sigma}$ is called the Standard normal variate (S.N.V)

It can be shown that the S.N.V Z has mean 0 and standard deviation 1.

Now,

$$\text{Mean of } Z = E(Z) = E = \left(\frac{X-\mu}{\sigma}\right)$$

$$= \frac{1}{\sigma} E(X-\mu) \qquad [\because E(cX) = cE(X)]$$

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

$$\text{Var }(Z) = \text{Var}\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma^2}\text{ Var }(X-\mu)$$

$$[\because \text{Var }(cX) = c^2 \text{ Var }(X)]$$

$$= \frac{1}{\sigma^2}\text{Var}(X)$$

$$\therefore \text{Var }(Z) = \frac{1}{\sigma^2} \cdot \sigma^2 = 1$$

$$\text{Standard Deviation} = \sqrt{1} = 1$$

Here the p.d.f. of the S.N.V Z is given by

$$\phi(Z)=\frac{1}{\sqrt{2\pi}}e^{\frac{-z^2}{2}}, \qquad -\infty < Z < \infty \text{ ....(2)}$$

[Taking $x = Z$, $\mu = 0$ and σ is 1 in the equation of the normal curve (i)]

STOP TO CONSIDER

(1) The area under the SNV curve between the two ordinates at $Z = 0$ and any positive value can be obtained from the table of area under the standard normal curve. It is to be noted that

(a) The area between the two ordinates at $Z = 0$ and $Z = a(> 0)$ is same as the area between the two ordinates at $Z = 0$ and $Z = -a$ under the curve.

(b) The area between the two ordinates at $Z = a$ and $Z = b$ $(a, b > 0)$ is the same as the area between the two ordinates at $Z = -a$ and $Z = -b$ under the curve.

(c) For solving any problem relating to Normal distribution, first of all, the normal variable X is to be transformed to the standard normal variable as follows :

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

where μ and σ are respectively the mean and standard deviation of the normal variable X. The need for doing so is that different normal distribution will have different means and standard deviations and consequently the normal curves will be different in terms of mean (average) and disperssion although Skewness and Kurtosis of all the normal curves will be the same each being equal to zero. But in case of standard normal curve, we get the same curve, (with mean = 0, S.D. = 1, Skewness = Kurtosis = 0). Corresponding to any normal curve and the various areas under the standard normal curve between the two ordinates at $Z = 0$ and any positive value has been calculated with the help of integral calculus. These areas can be obtained from the table of area under standard normal curve.

2.14 Area Under Standard Normal Curve :

A graph of the standard normal curve $Y = P(Z)$ is shown in the figure below. The percentage distribution of the area under the curve is also indicated in the figure.

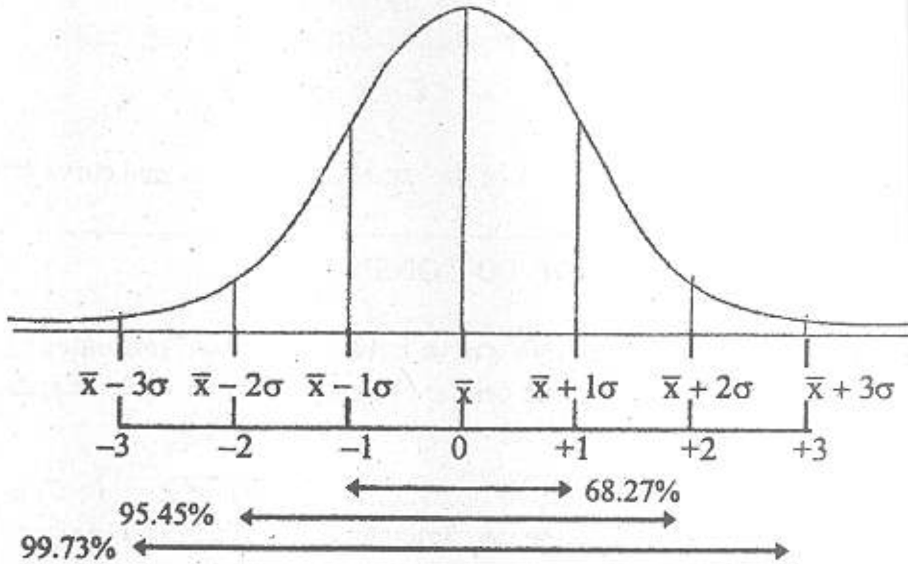


Figure : Area under normal curve and standard normal curve

The following table gives the area under the normal probability curve for some important values of standard normal variate Z.

Distance from the mean Ordinates in terms of $\pm \gamma$	Area under the curve
$Z = \pm 0.745$	50% = 0.50
$Z = \pm 1.00$	68.26% = 0.6826
$Z = \pm 1.96$	95% = 0.95
$Z = \pm 2.0$	95.44% = 0.9544
$Z = \pm 2.58$	99% = 0.99
$Z = \pm 3.0$	99.73% = 0.9973

SELF-ASKING QUESTION

Try yourself to mention 3 real life examples where normal distribution will be appropriate.

2.17 Solved Examples :

Example 1. The income of a group of 10,000 persons were found to be normally distributed. The mean income Rs. 520 and a standard deviation (S.D.) equal to Rs. 60.

(i) Find the number of persons having income between Rs. 400 and Rs. 550.

(ii) The lowest income of the richest 500.

[G.U. (MA/MSc)Prev'94]

Solution :

(i) Let X denotes the income of a person

Let, $X \sim N(\mu . \sigma)$

Given $\mu = 520, \sigma = 60, N = 10{,}000$

when $X = 400$, $Z = \frac{X-\mu}{\sigma} = \frac{400-520}{60} = -2$

when $X = 550$, $Z = \frac{X-\mu}{\lambda} = \frac{550-520}{60} = 0.5$

$$\begin{aligned} P(400 < X < 550) &= P(-2 < Z < 0.5) \\ &= P(-2 < Z < 0) + P(0 < Z < 0.5) \\ &= 0.4772 + 0.1915 \\ &= 0.5887 \end{aligned}$$

$\therefore$ Number of persons having income between 400 and 550 is,

$0.5887 \times 10{,}000 = 5887.$

(ii) Let, X_1 be the lowest income of the richest 500.

$P(X_1 \leq X \leq \infty) = \frac{500}{10{,}000} = 0.05$

when $X = X_1$, $Z = \frac{X_1 - 550}{60} = Z_1$ .....(i)

$P(X_1 \leq X \leq \infty) = 0.05$

$\Rightarrow P(Z_1 \leq Z \leq \infty) = 0.05$

$\Rightarrow 0.5 - P(0 < Z \leq Z_1) = 0.05$

$\Rightarrow P(0 < Z < Z_1) = 0.5 - 0.05 = 0.45$

$\therefore Z_1 = 1.65$

From (i), $\frac{X_1 - 520}{60} = 1.65$

$X_1 = 619$

$\therefore$ The lowest income of richest 500 persons were Rs. 619.

Example 2. Suppose the height of all cakes baked with a certain mix closely follow a normal distribution with mean 5.3 centemetre and SD 0.75 centemetre. (c.m).

(i) Find the percentage of cakes which have height of 4.4 c.m or less.

(ii) Number of such cakes in a lot of 800 cakes.

[G.U.(MA/MSc)Prev.'97]

Solution :

Let X denotes the height of a cake in c.m.

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

Given, $\mu = 5.3$, $\sigma = 0.75$

(i) When X = 4.4, $Z = \frac{X-\mu}{\sigma} = \frac{4.4-5.3}{0.75} = -1.2$

The percentage of cakes which have a height of 4.4 c.m or less w

be,

$$P(-\infty < X \le 4.4) = P(-\infty < Z \le -1.2)$$
$$= 0.5 - P(-1.2 \le Z \le 0)$$
$$= 0.5 - 0.3849$$
$$= 0.1151$$

(ii) Number of cakes in a lot of 800 cakes will be,

$$0.1151 \times 800 = 92.08 = 92 \text{ cakes.}$$

Example 3. In a certain book the frequency distribution for the numb of words per page may be taken as approximately normal with mean 80 and S.D. 50. If 3 pages are choosen at random, what is the probabili that none of them has both 830 and 845 words each?

[G.U. (MA/MSc)'02

Solution : Let X denotes the number of words in a page.

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

Given, $\mu = 800$, $\sigma = 50$

when X = 830, $Z = \frac{X-\mu}{\sigma} = \frac{830-800}{50} = 0.6$

when X = 845, $Z = \frac{X-\mu}{\sigma} = \frac{845-800}{50} = 0.9$

Now, $P(830 \le X \le 845) = P(0.6 \le Z \le 0.9)$

$$= P(0 < Z < 0.9) - P(0 < Z < 0.6)$$
$$= 0.3159 - 0.2257$$
$$= 0.0902$$

$\therefore$ Probability that none of them has between 830 and 845 wor each is,

$$1 - 0.0902 = 0.9098$$

Example 4. The income distribution of workers in a certain factory w found to be normal with a mean of rupees 500 and a S.D of Rs. 5 There were 228 persons getting above Rs. 600. How many perso were there in all? [G.U.(MA/MSc)'03]

Solution : Let, X denotes the income in rupees of a worker.

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

Given, $\mu = 500$, $\sigma = 50$, $N = ?$

$$\text{when } X = 600,\ Z = \frac{X-\mu}{\sigma} = \frac{600-500}{50} = 2$$

$$P(600 < X < \infty) = P(2 < Z < \infty)$$
$$= 0.5 - P(0 < Z < 2) = 0.5 - 0.4772 = 0.0228$$

Given,

$$\frac{228}{N} = 0.0228 \Rightarrow N = 10{,}000$$

Hence, there were 10,000 persons.

Example 5. A factory turns out an article by mass production method. From past experience it appears that 20 articles on an average are rejected out of every batch of 100. Find the S.D of the number of rejects in a batch and write the equation of the normal curve which may be taken to represent the distribution of the number of rejects in a large series of batches of 100. Hence, find the probability that the number of rejects in a batch exceeds 30. [G.U.(MA/MSc)'04]

Solution :

Let, X denotes the number of rejected articles in a batch.

i.e. $X \sim N(\mu, \sigma)$

$$\text{Given, } P = \frac{20}{100} = 0.2,\ n = 100$$

$$SD = \sqrt{npq} = \sqrt{100 \times 0.2 \times 0.8} = 4 \qquad \therefore \sigma = 4$$

The equation of the normal curve is,

$$f(x) = \frac{1}{4\sqrt{2\pi}} e^{\frac{-1}{2}\left(\frac{X-20}{4}\right)^2}, \qquad -\infty < X < \infty$$

$$\text{When, } X = 30,\ Z = \frac{X-\mu}{\sigma} = \frac{30-20}{4} = 2.5$$

$$P(30 < X < \infty) = P(2.5 < Z < \infty)$$
$$= 0.5 - P(0 < Z < 2.5)$$
$$= 0.5 - 0.4938 = 0.0062$$

Hence the probability that the number of rejects is a batch exceeds 30 is 0.0062.

CHECK YOUR PROGRESS

1. Draw a normal probability curve and describe its mai characteristics. [G.U.(MA/MSc)Prev '05,'08]
2. What is meant by normal distribution? What are its properties. [G.U.(MA/MSc)Prev '07,'10]
3. 1000 students took an examinations. The mean marks obtain is 55 and the standard deviation is 15. Find the number of students securing marks.

   (i) Between 40 to 80 (ii) Exceeding 70

   [G.U.(MA/MSc)Prev '91]
4. In a distribution exactly normal 7% of the items are under 32 and 85% are under 60. What are the mean and standard deviation of the distribution?
5. A set of examination marks is approximately normally distributed with a mean of 75 and a standard deviation of 5. If the top 5% of the students get grade A and bottom 25% get grade F. What mark is the lowest A and what mark is the highest F?

   [G.U.(MA/MSc)Prev '01]

2.16 Relation between Binomial and Normal Distribution :

Normal distribution is a limiting case of binomial probability distribution under the following conditions

(i) n, the number of trials is indefinitely large ie $n \to \infty$

(ii) Neither p nor q is very small.

De-Moivere proved that under the above two conditions, the distribution of standard binomial variate tends to the distribution of standard normal variate.

2.17 Relation between Poisson and Normal distribution :

If X is a random variable following poisson distribution with parameter m, then

$E(X) = \text{Mean} = m$ and

$\text{Var}(X) = \sigma^2 = m$

Thus, the standard poisson variate becomes

$$\frac{X - E(x)}{\sigma_x}$$

$$= \frac{X - m}{\sqrt{m}}$$

Hence, this variate tends to be standard normal variate if $m \to \infty$.

Comparison study of Binomial, Poisson and Normal

Properties	Binomial	Poisson	Normal
1. Nature	Discrete	Discrete	Continuous
2. Probability function	P(x) $= {}^nC_x P^x q^{n-x}$	P(x) $e^{-m} \cdot \frac{m^r}{r!}$	P(x) $= \frac{1}{r\sqrt{2\pi}} e^{\frac{-1}{2}\left(\frac{x-\mu}{r}\right)^2}$
3. Parameter Ristrictions	n, p $0 < p < 1$	m $m > 0$	$\overline{X} (= \mu), \sigma$ $-\infty \le X \le \infty$
4. Limiting form of distribution		$n \to \infty$ $p \to 0$ $np = m$ (finite)	$n \to \infty$ $p - q \to 0$ $p \to q \to \frac{1}{2}$
5. Mean and Variance	Mean = np Variance = npq	Mean = m Variance = m	Mean = μ Variance = σ^2
6. Shape	Symmetrical or Assymmetrical	Positively Skewed	Perfectly Symmetrical

2.18 Moment Generating Function :

The function $E\{e^{t(X-a)}\}$ serves to generate moments of the probability distribution of the variate about the point 'a' and is called the moment generating function (m.g.f) about 'a'. It is written as $M_{(x-a)}(t)$.

Here t is some arbitrary parameter.

Thus,

$M_{(x-a)}(t) = E[e^{t(x-a)}] = \sum e^{t(x-a)} p(x)$ ...... (i)

This is for discrete probability distribution with $P(x) = P(X = x)$.

For a continuous probability distribution having the density function f(x), the m.g.f. about some constant 'a' is defined as,

$M_{(x-a)}(t) = E[e^{t(x-a)}] = \int e^{t(x-a)} f(x)dx$ ...... (ii)

The summation of integration would extend to the entire range of the variate values.

By putting a = 0, we shall get m.g.f. about zero. It will generate moments about zero. Central moments m.g.f. about means is obtained by substituting m in place of a.

We shall show below how a m.g.f. generates moments for discrete distributions. Similarly we can show it for a continuous distribution.

$$M_{(x-a)}(t) = E[e^{t(x-a)}] = \sum_x e^{t(x-a)} p(x)$$

$$= \sum_x \left\{1 + t(x-a) + \frac{t^2(x-a)^2}{2!} + ... + \frac{t^r(x-a)^r}{r!} + ...\right\} p(x)$$

$$= \Sigma P(x) + t\Sigma(x-a)p(x) + \frac{t^2}{2!}\Sigma(x-a)^2 p(x) + ... +$$

$$\frac{t^r}{r!}\Sigma(x-a)^2 p(x) + ...$$

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

Or $M_{(x-a)}(t) = 1 + t\mu_1' + \frac{t^2}{2!}\mu_2' + ... + \frac{t^r}{r!}\mu_r' + ...$

where $\mu_r' = E(X-a)^r$ is the rth moment about the point x = a

From (i) and (ii) it is clear that,

$M_{(x-a)}(t) = E\{e^{t(x-a)}\}\ e^{-at} M_x(t) = e^{-at} \times$ m.g.f. about zero.

Thus, the m.g.f. about 'a' is equal to e^{-at} times the m.g.f. about zero.

STOP TO CONSIDER

Properties of M.G.F. :

(i) The m.g.f. of the sum of independent random variables is equal to the product of their respective m.g.f.

Symbolically, if $X_1, X_2, ... X_n$ be n independent random variables then the m.g.f. of their sum is given by,

$$M_{x_1+x_2+x_3+....+x_n}(t) = M_{x_1}(t) \times M_{x_2}(t) \times M_{x_3}(t) \times ..... \times M_{x_n}(t)$$

(ii) $M_{cx}(t) = M_x(ct)$, c being a constant.

2.19 Central Limit Theorem :

The central limit theorem may be stated as follows, "If $\overline{X}$ be the mean of a random sample of size n drawn from a population having mean μ and standard deviation σ, then the sampling distribution of the sample mean $\overline{X}$ is approximately a normal distribution with mean μ and standard deviation $\sigma / \sqrt{n}$, provided the sample size n is sufficiently large."

The central limit theorem tells us that, irrespective of the shape of the original distribution (i.e. whether the origin distribution is normal or not), the sampling distribution of the sample mean approaches the shape of a normal curve as the sample size becomes larger and larger. However, if the original population is distributed normally, the sampling distribution of the sample mean will also be normal whatever be the size of the sample.

Usually, a sample of 30 or more is considered as a 'large sample'. However, the larger the value of n, the better is the approximation.

The above central limit theorem is, in fact, a deduction from the following generalised central limit theorem :

"If $X_1, X_2 ...., X_n$ are independent random variables following any distribution, then under certain very conditions, their sum

$\sum X = X_1 + X_2 + ... + X_n$

is asymptotically normally distributed i.e.

$\sum X$ follows normal distribution as $n \rightarrow \infty$."

From the above generalised central limit theorem it has been proved that the sampling distribution of most statistics like sample proportion (p), difference of sample proportions $(p_1 - p_2)$, difference of sample means $(s_1 - s_2)$, difference of sample standard deviation etc are assymptotically normal, and thus the standardised variates corresponding to any one of these statistics is N(0, 1). Thus, if t is any statistic, then by central limit theorem,

$$Z = \frac{t - E(t)}{SE(t)} \sim N(0, 1)$$

assymptotically as $n \rightarrow \infty$. This result is extensively used in Large Sample Tests and also in constructing confidence limits for the population parameters when samples are large.

2.20 Summing Up :

Binomial distribution, also known as the 'Bernoulli Distributions' is the most fundamental and important discreet probability distribution and is defined by the probability function $P(x) = {}^nC_x p^x q^{n-x}$ where $x = 0,1,2,...n$. It gives the probability of X success in a series of an independent trials $(x \leq n)$, when P is the probability of success in a single trial and $q = 1 - p$. The binomial distribution is completely known when the values of its two parameters, 'n' and 'p' are known. The mean of the binomial distribution is np and variance is npq.

Poisson distribution is another discrete probability distribution and is applied when the number of trials is indefinitely large and the probability of success for each trial is indefinitely small. The Poisson distribution is defined by the probability function

$$P(X = r) = \frac{e^{-m}.m^r}{r!} \qquad r = 0,1,2,3,...$$

Poisson distribution is completely known when we know the value of its only parameter 'm'. The mean of Poisson distribution is 'm' and the variance is also 'm'.

Unlike binomial and Poisson distribution, normal distribution is a continuous probability distribution. A continuous distribution is one in which the underlying variable may assume any value within a range. The normal curve is a bell-shaped symmetric curve. Since, the curve is symmetric, hence, under normal distribution,

$$\text{mean} = \text{median} = \text{niode} = \mu$$

The function $E\{e^{t(x-a)}\}$ serves to generate moments of the probability distribution of the variable about the point 'a' and is called the moment generating functions about 'a'. It is written as $M_{(x-a)}(t)$

According to Central Limit Theorem; if $\overline{X}$ is the mean of a random sample of size n drawn from a population having mean μ and standard deviation σ, then the sampling distribution of the sample mean $\overline{X}$ is approximately a normal distribution with mean μ and standard deviation $\sigma / \sqrt{n}$ provided the sample size n is sufficiently large.

2.21 References and Suggested Readings :

1. Gupta, S.C. and Kapoor, U.K. "Fundamentals of Mathematical Statistics."
2. Gupta, S.C. and Kapoor, V.K., "Fundamentals of Applied Statistics."
3. Gupta, S.C., "Fundamentals of Statistics."
4. Agarwal, D.R., "Business Statistics."

●●●

Unit 3 : Income Distribution

Contents :

3.1 Introduction :

The issue of income distribution has always been a central concern of economic theory and economic policy. Earlier, the classical economists such as Adam Smith, Thomas Malthus and David Ricardo were mainly concerned with factor income distribution, which means the distribution of income between the main factors of production i.e. land, labour, capital. While modern economists have also addressed this issue, they havebeen more concerned with the distribution of income across individuals and households. Important theoretical and policy concerns include the relationship between income inequality and economic growth.

For households and individuals income is the sum of all wages, salaries, profits, interest payments, rents and other forms of earnings received in a given period of time.

3.2 Objectives :

This unit is designed to help you understand the pattern of income distribution in the economy. After reading this unit you will be able to,

1. Describe income and income distribution.

2. Distinguish between the absolute and relative measures of inco distribution.

3. Comment on the pattern of income distribution in the society i the world, country, state or in your locality.

4. Predict suitable distribution of income pattern in the econom

3.3 What is income :

Both of the above measures use income as the basis for evaluati poverty. However, 'income' is here understood in difference to common understanding. It means the total amount of goods and servic that a person receives, and thus there is not necessarily money cas involved. If a poor subsistence farmer grows his/her own grain it wi count as income. Services like public health and education are als counted in. Often expenditure or consumption (which is the same in th economic sense) is used to measure income.

The World Bank uses the so called living standard measureme surveys (LSMS) to measure income. These consists of questionnair with over 200 questions. Surveys have been completed in mos developing countries.

3.4 Income distribution in selected countries :

(a) United States of America :

In the United States, income was distributed somewhat inequall with those in the top two quintiles earning more tham the bottom 60% combined. Yet, the distribution of income was not as polarised as in man developing countries with most of America's earned income resting in th hands of the middle class. The following table illustrates the incom distribution of the United States for 2005.

Table : 4.1 : Income Distribution in USA (2005)

Income Range	Percentage of total income earned by the income group	Cummulative Percentage of Total Income
Less than 25,000	6.76	6.76
$ 25,000 to $ 50,000	18.12	24.88
$ 50,000 to $ 75,000	22.54	47.42
$ 75,000 to $ 100,000	20.00	67.42
$ 150,000 or more	32.58	100.00

(b) India :

In India, the income distribution as per the results of survey conducted in 1997 are tabulated below :

Table : 4.2 : Income Distribution in India (1997)

Income share held by	Percentage of total income earned by the income group	Cummulative Percentage of Total Income
Lowest 20% of the population	8.1	8.1
Second 20%	11.6	19.7
Third 20%	15.0	34.7
Fourth 20%	19.3	54.0
Highest 20%	46.0	100.0

It may be noted that the income share enjoyed by the highest 10% of the population of India was found to be as much as 33.5%, while the poorest 10% had to make do with a share of 3.5% only as per the above survey.

3.5 Pareto's law of Income distribution :

The Pareto distribution, named after the Italian economist Vilfredo Pareto, is a power law probability distribution found in a larger number of real-world situations. Outside the field of economics it is at times referred to as Bradford distribution.

Pareto originally used this distribution to describe the allocation of wealth among individuals since it seemed to show rather well the way that a larger portion of the wealth of any society is owned by a smaller percentage of the people in that society. This idea is sometimes expressed more simply as the Pareto principle or the "80-20 rule" which says that 20% of the population owns 80% of the wealth.

This distribution is not limited to describing wealth or income distribution, but to many situations in which an equilibrium is found in the distribution of the 'small' to 'large'. The following examples are sometimes seen as approximately Pareto distributed.

- Frequencies of words in longer texts (a few words are used often, lots of words are used infrequently).
- The sizes of human settlements (few cities, many hamlets/ villages).
- File size distribution of Internet traffic which uses the TCP protocol (many smaller files, few larger ones).
- The values of oil reserves in oil fields (a few large fields, many

small fields).

- The length distribution in jobs assigned supercomputers (a few larger ones, many smaller ones).

The Pareto law of income distribution states that 'The logarithm of the percentage of units with an income in excess of some value is a negatively sloped linear function of the logarithm of that value.

Symbolically, this takes the form,

$$\log P(y) = \log A - \alpha \log y$$

$$\Rightarrow P(y) = \frac{A}{Y^{\alpha}} \quad \text{where,}$$

P(y) = Percentage of units with an income in excess of y.

Y = Income level.

A, α = Parameters of the distribution. (α is called the Pareto Index).

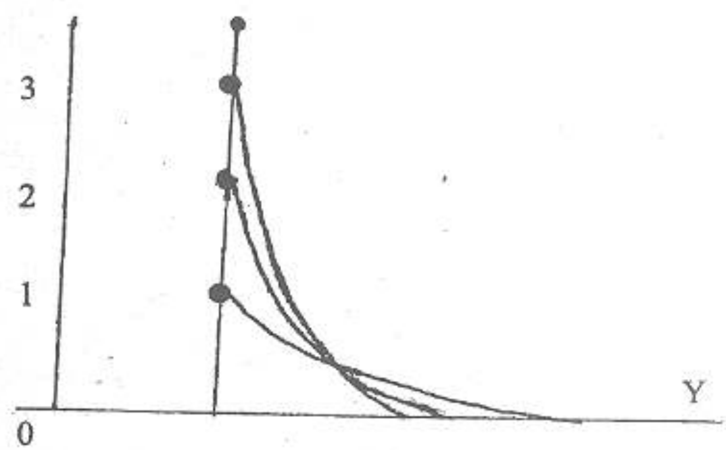
It may be noted that inequality varies inversely with 'α', which is the Pare to index.

The Pareto distribution is a cumulative distribution function. The corresponding density function of this distribution is given by

$$\Rightarrow P(y) = \alpha \frac{A^{\alpha}}{Y^{\alpha+1}}$$

As the income level approaches zero, the relative frequency p(0) in the above formula approaches infinity. On the other hand, as income gets larger and larger, the frequency falls towards zero.

The Pareto distribution is usually assumed to represent the distribution of income at upper levels. In case of income distributions, it does not fit into the distributions of low incomes well. In practice, those income units below the income tax levels donot fit well under the purview of the Pareto distribution.



The X-axis of the adjacent graph represents the income levels, with the initial income. $y_o = 1$ The percentage of units is represented on the Y-axis.

STOP TO CONSIDER

The following points may be noted for the Pareto distribution,

1. In order to determine whether a body of economic data follows the Pareto distribution, the data is to be plotted on a double logarithmic scale. If the resulting scatter of points lies along a negatively inclined straight line, it can be concluded that the data fits the Pareto distribution.

2. The Pareto distribution is a cumulative distribution. On the X-axis, the income level is represented, while the Y-axis represents the percentage of units with an income in excess of Y.

3. If the scale used on X-axis and Y-axis are not logerithmic, the Pareto distribution looks to as follows :

3.6 Log-normal Distribution :

The log normal distribution is the probability distribution of any random variable whose logarithm is normally distributed. If y is a random variable with a normal distribution, then e^y has a Log-normal distribution. A variable might be modeled as Log-normal if it can be thought of as the multiplicative product of many small independent factors. A random variable y is said to have the Lognormal distribution with parameters μ and σ if ln(y) has the normal distribution with mean μ and standard deviation σ. The log, normal distribution is used to model continuous random questions when the destribution is believed to be skewed to the right, such as certain income and lifetime variables. The probability density function starts at zero, increases to its mode and decreases thereafter.

Log-normal is also known as "log normal" or "lognormal." It is occasionally referred to as the Galton distribution or Galton's distribution after Francis Galton.

The base of the logarithmic function is immaterial since $\log_a y$ is normally distributed if and only if $\log_b y$ is normally distributed. It is assumed that only positive values of random variable y are considered.

The log-normal distribution has the density function given by the following formula.

$$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(\log y-\mu)^2}$$

for y (representing income) > 0.

Where μ and σ are the mean and standard deviation of y.

Further, since $\log y_2 - \log y_1 = \log (y_2/y_1)$; the differences in the logerithmic scale are a function of the ratios on the arithmetic scale.

To fit a log normal distribution to a body of data, we compute two sample statistics viz the arithmatic mean and the standard deviation of the data. The estimated mean and variance of the data is given by the following formulae.

$$\text{estimated } \mu = \left(\frac{1}{N}\right) \Sigma \log y_i$$

$$\text{estimated } \sigma^2 = \left(\frac{1}{N}\right) (\Sigma \log y_i - \text{estimated } \mu)^2$$

It may be noted that inequality varies directly with σ. Graphically, the cumulative log-normal function when plotted on a logarithmic scale should fall against a straight line.

STOP TO CONSIDER

Comments– Pareto and Log.normal distributions :

The Pareto and log-normal distributions are by and far, the most usual descriptions of income distributions. There is a view that Pareto distribution gives a better explanation at the upper tail, while log.normal distributions is better suited for lower income values.

This is due to following reason. Towards the greater end of a logarithmic table, given differences in arguments are associated with small differences in logarithm. Thus, the $\log_{10}1000$ is 3, while the $\log_{10}10000$ is only 4. On the other hand, at the smaller end, the same difference in arguments is associated with proportionately larger differences in logarithms. Thus, $\log_{10}1$ is 0, $\log_{10}1.5$ is 0.1761, $\log_{10}9$ is 0.3010 and $\log_{10}3$ is 0.4771.

The logarithmic scale thus compresses the distribution (of say income) at higher levels and stretches the same at lower level. Hence the merit of the log-normal distribution lies more in the lower values of income. But both the distribution of the exponential of random variables are distributed according to other common distribution respectively the exponential distribution and normal distribution.

3.7 Measurement of Income Distribution :

Generally, there are four criteria for inequality measurement. They are–

(1) Anonymity Principle :

From an ethical point of view, it does not matter who is earning the income. Permutations of incomes among people should not matter for inequality judgments.

$y_1 \leq y_2 \leq .... \leq y_n$.

This is the equivalent of arranging individuals so that they are ranked from poorest to richest (here y_i = income).

(2) Population Principle :

Cloning the entire population (and their incomes) should not alter inequality. The population principle is a way of saying that population size does not matter : all that matters are the proportions of the population that earn different levels of income.

The anonimity principle tells us that we can number people in order of increasing income and no useful information is lost. The population principle tells us that it does not matter how many people are there, we may normalise everything to percentages.

(3) Relative Income Principle :

This principle says that only relative incomes should matter and the absolute levels of these incomes should not. The advantage of this approach is that it enables us to compare income distributions for two countries that have different average income levels.

(4) The Dalton Principle :

The Dalton principle states that if one income distribution can be achieved from another by constructing a sequence of regressive transfers, then the new distribution must be deemed more unequal than the initial one.

Different techniques are used by economists to measure the distribution of income among members of a society. In particular, these techniques are used to measure the inequality, or equality of income within an economy. These techniques are typically catagorised as either absolute measures or relative measures.

3.8 Absolute Measures :

Absolute measures define a minimum standard, then calculate the number (or percent) of individuals below this threshold. These methods are most useful when determining the poverty in a society. Examples include,

- **Poverty line :**

This is a measure of the level of income necessary to subsist in a society and varies from place to place and from time to time depending on the cost of living and peoples' expectations. It is usually defined by governments and calculated as that level of income at which a household will devote two-thirds (to three quarters) of its income to basic necessities such as food, water, shelter and clothing.

- **Poverty Index :**

This index was developed by Amartya Sen. It takes into açcount both the number of poor and the extent of their poverty. Sen defined the index as,

$$I = (P/N) \times (B - A)/A$$

where, P = Number of people below the poverty line.

N = Total number of people in society.

B = Poverty line income.

A = Average income of those people below the poverty line.

3.9 Relative Measures :

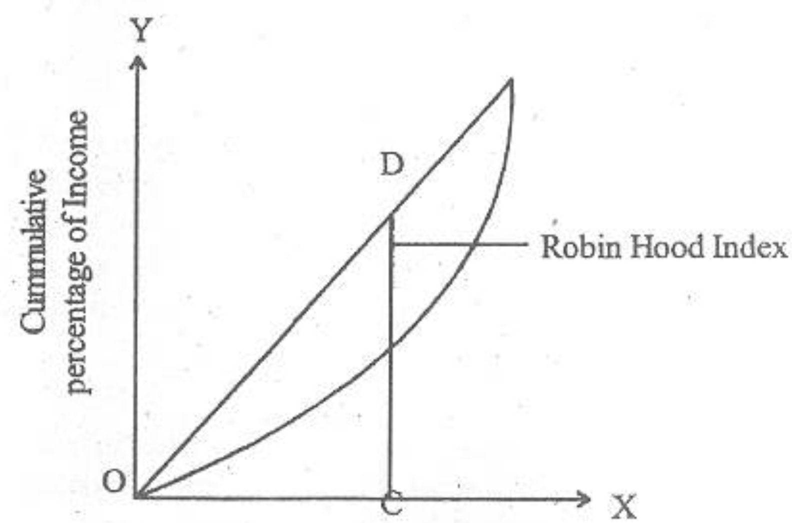
Relative income measures compare the income of one individual (or group) with the income of another individual (or group). These measures are most useful when analysing the scope and distribution of income inequality. Some examples include :

- **Percentile distribution :**

In this measure one percentile is compared to another percentile. For example, it might be determined that the income of the top ten-percentile is only slightly more than the bottom forty percentile. Or it might be determined that the top quartile earnes 45% of the society's income while the bottom quartile has 10% of society's income. The interquartile range in a standard percentile ranges from 25% to 75%.

- **Robin Hood Index :**

The Robin Hood index is a measurement of income inequality across a geographical area and is derived from the Lorenz curve. Mathematically, it is related to the Gini coeffecient, it measures the portion of total income that would have to be redistributed in order for there to be perfect equality. This index is derived by finding the largest vertical line, which can be drawn between a Lorenz curve for perfectly even distribution of incomes and the measured Lorenz curve. Robin Hood index is a measure of income inequality ranging from 0 (complete equality) to 100 (complete inequality). This index is also known as the Hoover index.



Cummulative percentage of households

- **Theil Index :**

This is a summary statistic used to measure income inequality, based on information entropy. It is similar to, but less commonly used than the Gini coeffecient. It has also been used to measure the lack of racial diversity.

- **Aitkinson Index :**

The Aitkinson index (also known as the Atkinson measure or Aitkinson inequality measure) is a measure of income inequality developed by Anthony Bames Aitkinson. Aitkinson index relies on the following axioms :

(i) The index is symmetric in its arguments.

(ii) The index is non-negative, and is equal to zero only if all incomes are the same.

(iii) This index satisfies the principle of transfers.

(iv) This index satisfies the population replication axiom; if a new population is formed by replicating the existing population an arbitary number of times, the inequality remains the same.

(v) The index satisfies mean independence, or income homogenity axiom : If all incomes are multiplied by a positive constant, the inequality remains the same.

- **Standard deviation of income :**

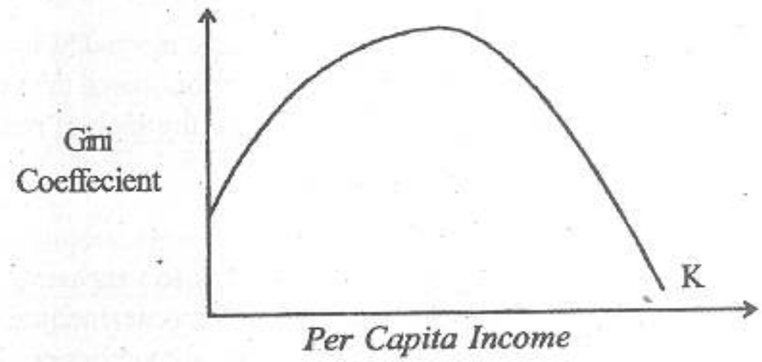
This measures income dispersion by assessing the squared variance from the mean. This metric is seldom seen, its use is limited to occasional reference in academic journals.

- **Relative Poverty line :**

This is a measure of the number or proportion of people or households whose level of income is less than some given fraction of typical incomes. This form of poverty measurement tends to concentrate concern on the bottom half of the income distribution and pay less attention to inequalities in the top half.

- **Kuznet Curve :**

Kuznet inverted-U shapped curve is a measure of income inequality. Kuznets suggested on the experience of the developed countries that historically there was a tendency for income inequality to increase first, and then to be reduced as countries developed from a low level.



There are other two most important measure of income inequality viz Lorenz curve and Gini Coeffecient.

CHECK YOUR PROGRESS

1. Discuss Pareto's law of Income Distribution. Compare and contrast it with the log normal distribution.
2. Discuss the criteria or principles for measuring inequality in income.
3. Discuss various absolute and relative measures used for measuring income inequality.
4. Write short notes on the following

   (a) Poverty Index (a) Robin Hood Index

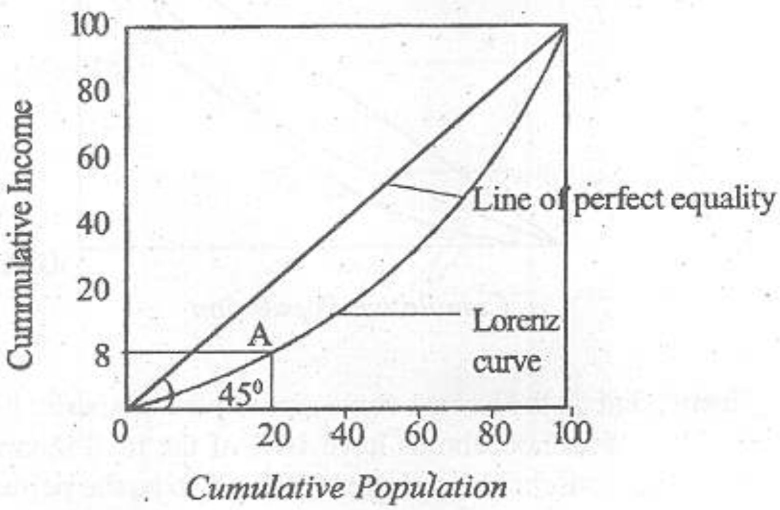
   (a) Aitkinson Index (a) Kuznet Curve

3.10 Lorenz Curve :

This is a graphic device used to display the relative inequality in a distribution of income values. A society's total income is ordered according to income level and the cumulative total graphed. Thus, the distribution of income within a community may be represented by the Lorenz Curve. The Lorenz curve is a graphical represenation of the cumulative distribution function of a probability distribution; it is a graph showing the proportion of the distribution assumed by the bottom N% of the values. On the horizontal axis, we depict cummulative percentage of the population arranged in increasing order of income. On the vertical axis, we measure the percentage of national income accuring to any particular fraction of the population thus arranged.

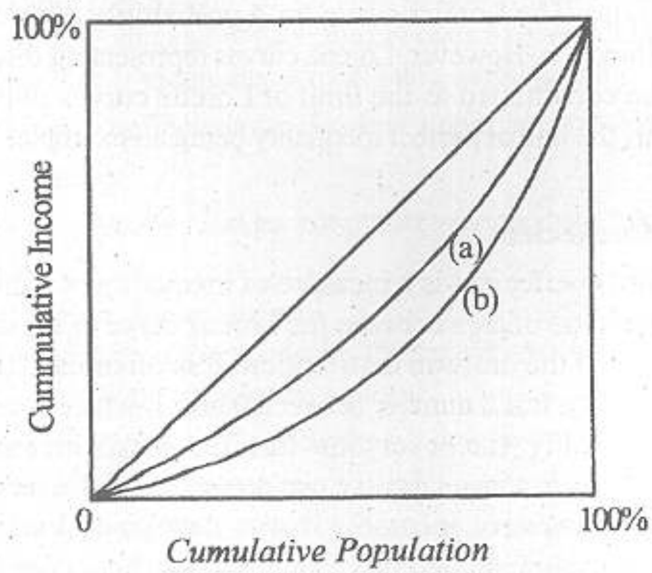
A perfectly equal income distribution would be one in which every person has the same income. In this case, the bottom N% of the society would always have N% of the income. This can be depicted by the straight line y = x; called the line of perfect equality or the 45⁰ line. Lorenz curve begins and ends on this 45⁰ line. This means that poorest 0% earn 0% of national income and the poorest 100% is just the whole population and so must earn 100% of the income. When everybody had the same income then Lorenz curve coincide everywhere with the 45⁰ line. But in practice, this type of instance is rare.

By contrast, a perfectly unequal distribution would be one in which one person has all the income and everyone else has none. In that case, the curve would be at $y = 0$ for all $x < 100\%$. If $y = 100\%$ when $x = 100\%$, this curve is called the line of perfect inequality.

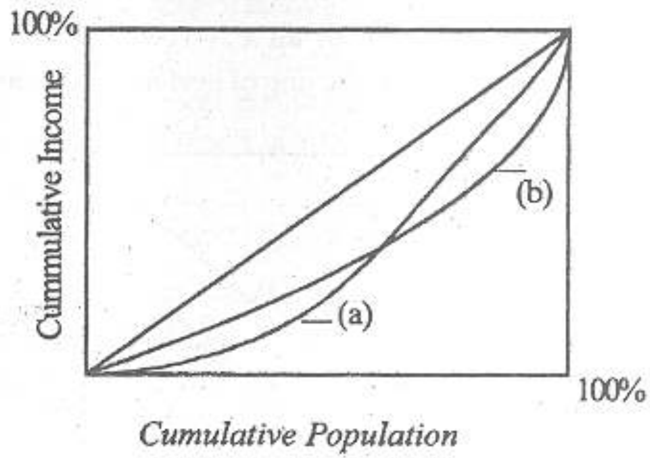


In the diagram above, poin A corresponds to a value of 20% on the population axis and 8% on the income axis. The interpretation is that the poorest 20% of the population earns only 8% of overall income.

The overall distance between the 45⁰ line and the Lorenz curve is indicative of the amount of inequality present in the society. The greater the extent of inequality, the further the Lorenz curve will be from the 45⁰ line.



In the diagram above, the curve (a) is always biased towards the poorest x% of the population, relative to (b). Hence (a) is more equal than (b). In this case we can compare the extent of inequality of the distribution of income. But sometime we cannot compare the inequality of the distribution on the basis of two Lorenz curves, when they intersect each other. In the diagram below we cannot say whether the Lorenz curve (a) or (b) is having more equal distribution or not, as neither Lorenz curve is uniformly to the right of the other.



Every point on the Lorenz curve represents a statement like, "The bottom 20% of all households have 10% of the total income." The percentage of households is plotted on the X-axis, the percentage of income on the Y-axis.

The Lorenz curve can also be used to show distribution of assets. It was developed by Max O Lorenz in 1905 for representing income distribution.

A Lorenz curve always starts at (0, 0) and ends at (1, 1). The Lorenz curve is not defined if the mean of the probability distribution is zero or infinite. The Lorenz curve for a probability distribution is a cumulative function. However, Lorenz curves representing discontinuous functions be constructed as the limit of Lorenz curves of probability distributions, the line of perfect inequality being an example.

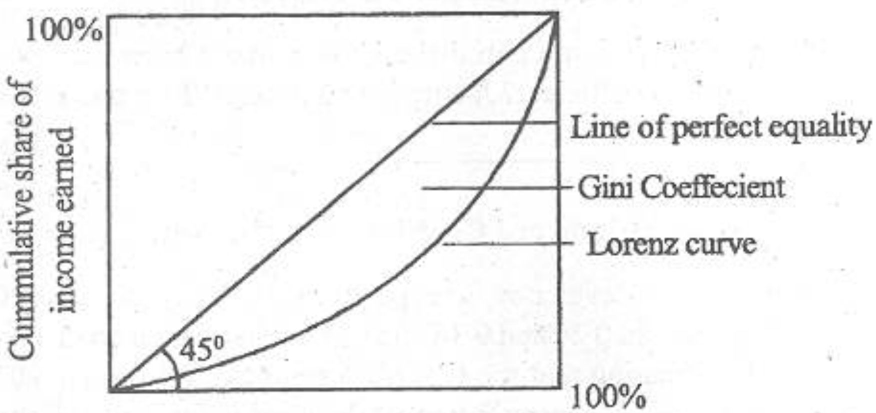
3.11 Gini Coeffecient :

The Gini coeffecient is a measure of inequality of a distribution, defined as the ratio of area between the Lorenz curve of the distribution and the curve of the uniform distribution. It is often used to measure income inequality. It is a number between 0 and 1, where 0 corresponds to perfect equality (i.e. everyone has the same income) and 1 corresponds to perfect inequality (i.e one person has all the income, while everyone else has zero income). It was developed by the Itallian statistician corrado Gini and published in his paper "Variability and

Mutability" (1912).

The Gini-coeffecient is equal to half of the relative mean difference. The Gini index is the Gini coeffecient expressed as a percentage, and is equal to the Gini coeffecient multiplied by 100. While the Gini coeffecient is mostly used to measure income inequality, it can also be used to measure wealth inequality. This use requires that no one has a negative wealth.

The Gini coeffecient is defined as a ratio of the areas on the Lorenz curve diagram. If the area between the line of perfect equality and Lorenz curve is A_1 and the area under the Lorenz curve is B, then the Gini coeffecient is $A/(A + B)$. Since $A + B = 0.5$, the Gini coeffecient, $G = 2A = 1 - 2B$.



The cumulative share of people from lower income

STOP TO CONSIDER

Advantages of Gini Coeffecient :

(1) Gini-coeffecient can be used to compare income distributions across different population sections as well as countries.

(2) Gini coeffecient can be used to indicate how the distribution of income has changed with a country over a period of time, thus it is possible to see if inequality is increasing or decreasing.

(3) Gini Coeffecient satisfies four important principles.

They are :

(a) **Anonymity :** It does not matter who the high and low earners are

(b) **Scale Independence :** The Gini coeffecient does not consider the size of the economy, the way it is measured. or whether it is a rich or poor country on average.

(c) **Population Independence :** It does not matter how large the population of the country is.

(d) **Transfer Principle :** If income is transferred from a rich person to a poor person, the resulting distribution is more equal.

Disadvantage of Gini Coeffecient :

(1) While the Gini coeffecient measures inequality of income, it does not measure inequality of opportunity.

(2) If two countries have the same Gini coeffecient but one is rich and the other is poor, it can be seen to measure two different things. In a poor country, it measures the inequality in material life quality while in a rich country it measures the distribution of luxury beyond the basic necessities.

(3) Gini coeffecient of different sets of people cannot be averaged to obtain the Gini coeffecient of all people in the sets.

SELF-ASKING QUESTION

Do you think you can make a comparison between Lorenz curve and Gini coefficient? Justify your answer. Which one is better according to you?

3.12 Income Gini Indices / Coeffecient in the World :

While most developed European nations tend to have Gini coeffecient between 0.24 and 0.36, the United States Gini coeffecient is above 0.4, indicating that the United States have greater inequality. According to the US census Bureau, Gini indices for United States at various times are :

2000 : 46.2	2005 : 46.9
2006 : 47.0 (highest index reported)	2007 : 46.3
2008 : 46.69	2009 : 46.8

Using the Gini coeffecient can help quantity differences in welfare and compensation policies and philosophies. However it should be borne in mind that the Gini coeffecient can be misleading when used to make political comparisons between large and small countries.

Poor countries (those with low per capita GDP) have Gini coeffecient that have over the whole range from low (0.25) to high (0.71), while rich countries have generally low Gini coeffecient (under 0.40).

3.13 Income Gini Indices / Coeffecients of Assam and other states of India :

While the overall income growth rate is slow, the Assamese society has been more egalitarian than most other Indian states. The Gini coeffecient in consumption expenditure distribution has been only around 0.19 for rural areas and around 0.29 for urban areas of Assam in recent years.

Among the major Indian states, Assam's income distribution is the most egalitarian in rural areas and the second best in urban areas on an average basis during 1990-94. The relevant details are tabulated below.

Table 4.3 : Gini Indices/Coeffecient of some Indian states (1990-91 to 1993-94)

State	Rural Area	Urban Area
Andhra Pradesh	28.39	32.50
Assam	19.27	28.94
Bihar	22.36	31.72
Gujrat	24.07	29.52
Jammu and Kashmir	27.87	28.45
Karnataka	34.09	34.63
Kerala	30.62	37.16
Madhya Pradesh	38.17	33.76
Maharastra	37.47	34.86
Orissa	43.31	37.83
Rajasthan	27.98	29.61
Tamil Nadu	29.39	36.82
Uttar Pradesh	28.09	32.75
West Bengal	25.75	34.37

Source : G. Datt, Indian Journal of Labor Economics, 1998.

3.14 Summing Up

Income means the total amount of goods and services that a person receives, and thus there is not necessarily money cash involved. Here wee have studied two types of income distributions viz., pareto distributions and log-normal distributions. Pareto distribution was originally used to describe the allocation of wealth among individuals since it seemed to show that a larger portion of the wealth of any society is owned by a smaller percentage of people. This idea is also expressed as the 'Pareto principle' or the '80-20 rule' which says 20% of the population owns 80% of the wealth.

The log-normal distribution is the probability distribution of any random variable whose logarithm is normally distributed.

There are 4 criteria for measuring inequality in income, viz., anonymity principle, population principle, relative income principle and the Dalton principle.

There are two types of measures to measure income inequality- Absolute measures and relative measures., Absolute measures define a minimum standard, then calculate the number (or percent) of individuals below this standard. For eg., poverty line, poverty index etc. Relative measures compare the income of one individual (or group) with the income of another individual (or group). For eg., percentile distribution, Robin Hood Index, Theil Index, Aitkinson Index etc.

CHECK YOUR PROGRESS

1. What is Lorenz Curve? Draw a Lorenz Curve.
2. Define Gini Coeffecient.
3. How Gini Coeffecient can be determined on the basis of Lorenz Curve.
4. Write a note on the merits and demerits of Gini coefficient.

3.15 References and Suggested Readings :

1. Salvatore, Dominick and Reagle, Darrick, "Statistics and Econometrics", TMH.
2. Klein, L.R., "An introduction to Econometrics."
3. Ray, Debraj, "Development Economics", Oxford University Press.
4. Hooda, P.R., "Statistics for Business and Economics", McMillan.

●●●

Unit 4 : Index Number

Contents :

4.1 Introduction :

An index number is a measure designed to show average changes in the price, quantity or value of a group of items with respect to time, geographical location or situation.

Index numbers are regarded as "barometers of economic activity." They help in framing suitable policies making forecast of future economic activity. Index numbers may be classified in terms of what they measure. In Economics and business the classification are :

(i) Price, (ii) Quantity, (iii) Value and (iv) Special Purpose.

The index numbers are indicators which reflect the relative changes in the level of certain phenomenon in any given period (or over a specified period of time) called the current period with respect to its value in some fixed period, called the base perise selected for comparison.

4.2 Objectives :

This unit is designed to help you understand the concept of index numbers. After reading this unit you will be able to,

1. Compare between Laspeyres' and Paasche's index number.

2. Construct shifting, splicing and deflating index number.

3. Predict the trend of economic fluctuation.

4. Formulate future policies on economics depending on the fluctuating trend.

4.3 System of Weighting in Index Numbers :

The commodities included for the construction of index numbers (like food, clothing, housing, light and fuel) are not of equal importance. In order that the index is representative of the average changes in the level of phenomenon for group, proper weights should be assigned to different commodities according to their respective importance in the group. If we ignore weights, the result is not simply an un-weighted index, but rather an inappropriate weighted index.

The weighted system adopted should truely reflect the importance of each commodity. Since an index would not depend on the units in which the price or quantities are reported, the prices are weighted by quantities, the quantities by prices while the price relatives are weighted by values. The prices and quantities used as weights may relate to the base period or to the current period.

Let us consider some important weighted price index numbers. The weighted quantity indices can be constructed in a similar manner.

If P_{oi} denotes the price of the i th commodity in the base period, P_{1i}, the price of the ith commodity in the certain period and W_i is the weight attached to the price relative for the ith commodity then weighted aggregate index is :

$P_{01} = \frac{\Sigma P_{1i} W_i}{\Sigma P_{0i} W_i}$. Again,

Arithmetic Mean	$P_{01} = \frac{\sum_i \frac{P_{1i}}{P_{0i}} Wi}{\sum_i W_i}$
Weighted Geometric Mean	$P_{01} = \left\{ \prod_i \left(\frac{P_{1i}}{P_{0i}} \right)^{wi} \right\}^{\frac{1}{\sum_i W_i}}$
Weighted Harmonic Mean	$P_{01} = \frac{\sum_i W_i}{\sum_i \frac{P_{0i}}{P_{1i}} W_i}$

Different statisticians has constructed the formula for constructing index number differently. They have assigned different weights for computations of index numbers. Some of the formula are given below :

Laspeyre's	Weight = q_{oi} = Base period quantities	$P_{01}^{La} = \frac{\Sigma P_{1i} q_{oi}}{\Sigma P_{oi} q_{0i}}$
Paasche	Weight = q_{1i} = Current period quantities	$P_{01}^{Pa} = \frac{\Sigma P_{1i} q_{1i}}{\Sigma P_{oi} q_{1i}}$
Marshall Edgeworth	Weight = $\frac{q_{1i} + q_{0i}}{2}$ = Average of current and base period quantities	$P_{01}^{ME} = \frac{\Sigma P_{1i}(q_{1i} + q_{0i})}{\Sigma P_{oi}(q_{1i} + q_{oi})}$
Irving Fisher	It is the geometric mean of Laspeyre's and Paasche's price index numbers.	$P_{01}^{F} = \sqrt{P_{01}^{La} \times P_{01}^{Pa}}$ $= \sqrt{\frac{\Sigma P_{1i} q_{oi}}{\Sigma P_{0i} q_{oi}} \times \frac{\Sigma P_{1i} q_{1i}}{\Sigma P_{0i} q_{1i}}}$
Durbish and Bowley	It is the arithmetic mean of Laspeyre's and Paasche's indices.	$P_{01} = \frac{L + P}{2}$ $= \frac{\frac{\Sigma P_{1i} q_{0i}}{\Sigma P_{0i} + q_{0i}} + \frac{\Sigma P_{1i} q_{1i}}{\Sigma P_{0i} q_{1i}}}{2}$
Kally	Here weights are quantities which may refer to some period, not necessarily the base year or current year.	$P_{01} = \frac{\Sigma P_{1i} q_i}{\Sigma P_{0i} q_i}$ where, $q_i = \frac{q_{0i} + q_{1i}}{2}$

Let us understand,

Types of Indices,

- Unweighted Index
  - Simple Average of the price Index
  - Simple Aggregate Index
- Weighted Index
  - Laspeyre's Price Index
  - Paasche's Price Index
- Fisher's Price Index

- Value Index
- Special Purpuse Index
  - – Consumer Price Index
  - – Producer Price Index
  - – S. & P Index

4.4 Relation between Laspeyre's and Paasche's Price Index Numbers :

Laspeyre's and Paasche's price index numbers are two most important numbers in the statistical and economic theory of index numbers. These two indices will in general, gives different results when applied to the same set of data since they measure the change in the cost of two distinct collection of commodities. These two indices will be equal if either price of all the goods (included in the index) change in the same ratio or if the quantities of the goods change in the same ratio. Since under normal circumstances, neither of these two indices is governed by the correlation between price and quantity movements, as a result, Laspeyre's index will show greater rise, when prices in general rise, when prices in general are rising and a smaller fall in a period of falling prices.

The exact relationship between the two indices can be derived by considering the formula for the coeffecient of linear correlation. If we have a series of pairs of observation of X and Y, each pair being weighted by a frequency f, the sum of the frequencies being N, the coeffecient of linear combination between X and Y is given by,

$$r_{XY} = \frac{\text{Cov}(X,Y)}{\sigma_x \sigma_y}$$

$$= \frac{\frac{1}{N}\Sigma f(X-\overline{X})(Y-\overline{Y})^2}{\sqrt{\frac{1}{N}\Sigma(X-\overline{X})^2}\sqrt{\frac{1}{N}\Sigma(Y-\overline{Y})^2}}$$

$$= \frac{\frac{\Sigma f(XY)}{N} - \frac{\Sigma fX}{N}\frac{\Sigma fY}{N}}{\sigma_x \cdot \sigma_y}$$

In the above expression, σ_x and σ_y are the standard deviations of the X and Y series respectively with the f's included as weights.

We substitute $\frac{p_{1i}}{p_{0i}}$ for X, $\frac{q_{1i}}{q_{0i}}$ for Y and $P_{oi}\, q_{oi}$ for f in the above expression. Upon substitution we have,

$$r_{xy} = \frac{\dfrac{\sum p_{0i} - q_{oi}\left(\dfrac{p_{1i}}{p_{oi}}\right)\left(\dfrac{q_{1i}}{q_{oi}}\right)}{N} - \dfrac{\sum p_{0i}.q_{oi}\left(\dfrac{p_{1i}}{p_{oi}}\right)}{N}.\dfrac{\sum p_{0i}.q_{oi}\left(\dfrac{q_{1i}}{q_{0i}}\right)}{N}}{\sigma_X.\sigma_Y}$$

N being the sum of frequencies, $f(\text{i.e.,} \sum p_{0i}q_{oi})$

$$r_{xy}.\sigma_X.\sigma_Y = \frac{\sum\left(p_{oi}q_{oi}.\dfrac{p_{1i}q_{1i}}{p_{oi}q_{oi}}\right)}{\sum p_{oi}q_{oi}} - \left[\frac{\sum\left(p_{oi}q_{oi}.\dfrac{p_{1i}}{p_{oi}}\right)}{\sum p_{oi}q_{oi}}.\frac{\sum\left(p_{oi}q_{oi}.\dfrac{q_{1i}}{q_{oi}}\right)}{\sum p_{oi}q_{oi}}\right]$$

$$\Rightarrow r_{XY}.\sigma_X.\sigma_Y = \frac{\sum p_{1i}q_{1i}}{\sum p_{oi}q_{oi}} - \left(\frac{\sum p_{1i}q_{oi}}{\sum p_{oi}q_{oi}}\right)\left(\frac{\sum p_{oi}q_{1i}}{\sum p_{oi}q_{oi}}\right)$$

$$\Rightarrow \left(\frac{\sum p_{1i}q_{oi}}{\sum p_{oi}q_{oi}}\right)\left(\frac{\sum p_{oi}q_{1i}}{\sum p_{oi}q_{oi}}\right) = \frac{\sum p_{1i}q_{1i}}{\sum p_{oi}q_{oi}} - r_{XY}.\sigma_X.\sigma_Y$$

if $\dfrac{\sum p_{1i}q_{1i}}{\sum p_{oi}q_{oi}}$ is defined as V_{oi}, which is the index of value expanded between the base period and the *i* th period, then,

$$\left(\frac{\sum p_{1i}q_{oi}}{\sum p_{oi}q_{oi}}\right)\left(\frac{\sum p_{oi}q_{1i}}{\sum p_{oi}q_{oi}}\right) = V_{oi} - r_{XY}.\sigma_X.\sigma_Y$$

dividing both sides by $\dfrac{\sum p_{1i}q_{1i}}{\sum p_{oi}q_{oi}}$ (or V_{oi}) we get,

$$\frac{\left(\dfrac{\sum p_{1i}q_{oi}}{\sum p_{oi}q_{oi}}\right)\left(\dfrac{\sum p_{oi}q_{1i}}{\sum p_{oi}q_{oi}}\right)}{\left(\dfrac{\sum p_{1i}q_{1i}}{\sum p_{oi}q_{oi}}\right)} = 1 - \frac{r_{XY}.\sigma_X.\sigma_Y}{V_{oi}}$$

$$\Rightarrow \left(\frac{\sum p_{1i}q_{oi}}{\sum p_{oi}q_{oi}}\right)\left(\frac{\sum p_{oi}q_{1i}}{\sum p_{1i}q_{oi}}\right) = 1 - \frac{r_{XY}.\sigma_X.\sigma_Y}{V_{oi}}$$

$$= P_{o1}^{La} \times \frac{1}{P_{o1}^{Pa}} = 1 - \frac{r_{XY}.\sigma_X.\sigma_Y}{V_{oi}}$$

$$= \frac{P_{o1}^{La}}{P_{o1}^{Pa}} = 1 - \frac{r_{XY}.\sigma_X.\sigma_Y}{V_{oi}}$$

From the above expression it is clear that the expression on the left hand side, ie $\frac{P_{01}^{La}}{P_{01}^{pa}}$ would be equal to unity and consequently $P_{01}^{La} = P_{01}^{pa}$, only if either r_{xy}, σ_x or σ_y is equal to zero.

Normally, since

$$-1 < r_{xy} < 1 \text{ and } \sigma_x \neq 0 \text{ and } \sigma_y \neq 0$$

$$\therefore\ P_{01}^{La} > P_{01}^{pa}.$$

Paasche's index has a downward bias while Laspeyre's index has an upward bias. Paasche's index in comparison to Laspeyre's index, shows a smaller rise when the prices in general are rising and a greater fall when the prices in general are falling. This can be expressed by saying that Laspeyre's price index tends underestimate price changes. However, Laspeyre's price index is not always higher than Paasche's price index.

STOP TO CONSIDER

Laaspeyre's Index :

Advantage :

It requires quantity data from only the base period. This allows a more meaningful comparison over time. The changes in the index can be attributed to changes in the price.

Disadvantage :

It does not reflect changes in buying patterns over time. Also it may over weight goods whose prices increase.

Paasche's Index :

Advantage :

It uses quantities from the current period. Hence it reflects current buying habits.

Disadvantage :

It requires quantity data for the current year. As different quantities are used each year, it is impossible to attribute changes in the index to changes in the price alone. It tends to overweight the goods whose prices have declined. It requires the prices to be recomputed each year.

4.5 Test for Index Numbers :

According to Irving Fisher, any good index number should satisfy two sets of tests viz. the time reversal test and factor reversal test.

4.5.1 Time Reversal Test :

Time reversal test is a test to determine whether a given method will work both ways in time forward and backward. In the words of Irving Fisher, "The test is that the formula for calculating an index number should be such that it will give the same ratio between one point of comparison and the other, no matter which of the two is taken as base or putting in other way, the index number reckoned forward should be reciprocal of that reckoned backward."

Time reversal test is satisfied by the following index number formulae.

(1) Simple Aggregate Index.

(2) Marshall Edgeworth Formula.

(3) Walsch Formula.

(4) Fisher's Ideal Formula.

(5) Kelly's Fixed Weight Formula.

(6) Simple Geometric Mean or Price Relatives Formula.

(7) Weighted Geometric Mean of Price Relatives Formula with Fixed Weights.

Laspeyre's and Paasche's index numbers donot satisfy the time reversal test.

Proof : Laspeyre's price index is,

$$P_{01} = \frac{\Sigma P_1 q_0}{\Sigma P_0 q_0}$$

$$\therefore P_{10} = \frac{\Sigma P_0 q_1}{\Sigma P_1 q_1}$$

$$\text{Now, } P_{01} \times P_{10} = \frac{\Sigma P_1 q_0}{\Sigma P_0 q_0} \times \frac{\Sigma P_0 q_1}{\Sigma P_1 q_1} \neq 1$$

Paasche's price index is,

$$P_{01} = \frac{\Sigma P_1 q_1}{\Sigma P_0 q_1}$$

$$\therefore P_{10} = \frac{\Sigma P_0 q_0}{\Sigma P_1 q_0}$$

$$\text{Now, } P_{01} \times P_{10} = \frac{\Sigma P_1 q_1}{\Sigma P_0 q_1} \times \frac{\Sigma P_0 q_0}{\Sigma P_1 q_0} \neq 1$$

But if Laspeyre's price index is equal to Paasche's price index, then both of these index numbers satisfy the time reversal test.

According to this test, any index number formula to be accurate should be time consistant. Thus, P_{01} measure price movements between period 0 (base period) and period 1 (current period), while P_{10} measures the price movements between period 1 (base period) and period 0 (current period), we ought to have,

$$P_{01} = \frac{1}{P_{10}}$$

$$\Rightarrow P_{01}\,P_{10} = 1$$

The left hand side of the equations, (i)

$$\frac{\Sigma P_{1i} q_{0i}}{\Sigma P_{0i} q_{0i}} \frac{\Sigma P_{0i} q_{1i}}{\Sigma P_{1i} q_{1i}} = 1 - \frac{r_{xy} \sigma_x \sigma_y}{\gamma_{01}}$$

Can be expressed as,

$$P_{01}^{La} \Big/ P_{01}^{Pa} \qquad \text{Or} \qquad \frac{1}{P_{01}^{pa} \Big/ P_{01}^{La}}$$

Hence, neither Laspeyre's nor Paasche's index number formula will satisfy the time reversal test except when either r_{xy} or σ_x or σ_y equals to zero, i.e. the conditions which are rarely fulfilled under normal circumstances. Besides, normally Laspeyre's index shows an upward bias while Paasche's index shows a downward bias of the same relative magnitude in relation to the test. The test is satisfied by the simple geometric mean of price relatives the simple aggregative index, by the median and mode of the price relatives as well as by the Edgeworth Marshall formula and Fisher's ideal index number. The weighted geometric mean of price relatives and the weighted aggregative index will satisfy the time reversal test if the weights are constant.

Fisher's ideal formula satisfy the test.

Proof :

$$P_{01} = \sqrt{\frac{\Sigma P_1 q_0}{\Sigma P_0 q_0} \times \frac{\Sigma P_1 q_1}{\Sigma P_0 q_1}}$$

Changing time i.e. 0 to 1 and 1 to 0,

$$P_{10} = \sqrt{\frac{\Sigma P_0 q_1}{\Sigma P_1 q_1} \times \frac{\Sigma P_0 q_0}{\Sigma P_1 q_0}}$$

$$P_{01}P_{10} = \sqrt{\frac{\Sigma P_1 q_0}{\Sigma P_0 q_0} \times \frac{\Sigma P_1 q_1}{\Sigma P_0 q_1} \times \frac{\Sigma P_0 q_1}{\Sigma P_1 q_1} \times \frac{\Sigma P_0 q_0}{\Sigma P_1 q_0}}$$

$$= \sqrt{1} = 1$$

Hence, Fisher's index number satisfy the time reversal test.

4.5.2 Factor Reversal Test :

Factor Reversal test holds that the product of a price index and the quantity index should be usual to the corresponding value index. In the word of fisher, "Just as each formula should permit the interchange of the two times without giving inconsistent results, so it ought to permit interchanging the prices and quantities without giving inconsistent result. i.e. the two results multiplied together should give the true value ratio.

In the test the change in price multiplied by the change in quantity should be equal to the total change in value. The total value of a given commodity in the price per unit (total value = p × q). If p_1 and p_0 represent prices and q_1 and q_0 represent quantities in the current year and base year respectively, and if p_{01} represents the change in price in current year and Q_{01} is the change in quantity in the current year, then

$$P_{01} \times Q_{01} = \frac{\Sigma P_1 q_1}{\Sigma p_0 q_0}$$

The factor reversal test is satisfied only by Fisher's ideal index number.

Proof :

$$P_{01} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}}$$

Changing p to q and q to p we have,

$$Q_{01} = \sqrt{\frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} \times \frac{\Sigma q_1 p_1}{\Sigma q_0 p_1}}$$

$$P_{01} \times Q_{01} = \sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_1 q_1}{\Sigma p_0 q_1} \times \frac{\Sigma q_1 p_0}{\Sigma q_0 p_0} \times \frac{\Sigma q_1 p_1}{\Sigma q_0 p_1}}$$

However Fisher's index number, although satisfies the above mentioned two tests suffer from lack of clarity of meaning. It is not at clear, the change in which collection of goods is measured by Fisher's ideal index number.

4.5.3 Chained Indices : The Circular Test :

We have considered so far that base period remains fixed. However, with the passage of time, the various types, quality and importance of the various commodities as well as the tastes and habits of the consumers undergo a change. Thus, if we wish to compare the price movements over successive time periods 0, 1, 2, ..., n instead of using a fixed base period 0, we compute number of link indices $p_{01}, p_{12}, \ldots, p_{n-1}, p_n$. Thus by chaining together the link indices we obtain the chained indices as follows–

$$P_{01}$$

$$P_{02}^{ch} = P_{01}P_{12}$$

$$P_{03}^{ch} = P_{01}P_{12}P_{23}$$

$$. = .$$

$$P_{0n}^{ch} = P_{01}P_{12}.....P_{n-1}P_n$$

Now any chained index P_{0n}^{ch} will be equal to the coresponding fixed base index if it satisfies the circular test which can be stated symbolically as,

$$P_{01}P_{12}.....P_{n-1}P_nP_{no} = 1$$

The circular test is an extension of time reversal test for more than two periods and is based on the shiftability of base period. The circular test is not met by simple geometric mean of price relatives and the weighted aggregative fixed weights. Laspeyre's and Paasche's index numbers and their derivatives, the Marshall-Edgeworth and the Ideal indices donot meet the circular test. Because, the weights in those index numbers depend on the periods between which comparisons are being made. If these periods change, the weight change.

When the test is applied to simple aggregative method, we will get,

$$\frac{\Sigma p_1}{\Sigma p_0} \times \frac{\Sigma p_2}{\Sigma p_1} \times \frac{\Sigma p_0}{\Sigma p_2} = 1$$

Similarly, when applied to fixed weight aggregative method, we get,

$$\frac{\Sigma p_1 q}{\Sigma p_0 q} \times \frac{\Sigma p_2 q}{\Sigma p_1 q} \times \frac{\Sigma p_0 q}{\Sigma p_2 q} = 1$$

Laspeyre's index does not satisfy this test.

$$P_{01}^{La} \times P_{12}^{La} \times P_{20}^{La} = \frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_2 q_1}{\Sigma p_1 q_1} \times \frac{\Sigma p_0 q_2}{\Sigma p_2 q_2} \neq 1$$

STOP TO CONSIDER

Uses of Chain Base Index Number :

(i) In the chain base method the comparison are made with the immediate past (preceding year) and accordingly the data (for the two periods being compared) are relatively homogenous. The comparisons are therefore, more valid and meaningful and resulting index is more representative of the current trends in the tastes, habits, customs and fashions of the society.

(ii) In this method, new commodities or items may be included and old and obsolute items may be deducted without impairing comparability and without requiring the recalculation of the entire series of index numbers, which is necessary in case of fixed base method.

Differences between Fixed Base and Chained Base Methods

Fixed Base	**Chain Base**
1.It has a fixed base	1.It's base period changes
2.When the base year is from a distant past, it becomes irrelevant.	2.Comparison becomes relevant.
3.In fixed base, base year is always a normal year.	3.Here, as the base year is the preceding period, sometimes that year may not be a normal year.In that case the result may be doubtful.

4.6 Fisher's Index Number :

Laspeyre's index number tends to overweight goods whose price have increased. Paasche's index tends to overweight goods whose price have gone down. Fisher's ideal index was developed in an attempt to offset these shortcomings.

Fisher's index number is called the ideal index number, because,

(1) It is based on the geometric mean which is theoretically considered to be the best average for constructing index numbers.

(2) It takes into account both current year as well as base year prices and quantities.

(3) It satisfies both the time reversal test as well as the factor reversal test.

(4) It is free from any bias.

Example 1 :

Prove that Fisher's ideal index number lies between Laspeye's and Paasche's index numbers.

Solution :

Let us consider two real positive numbers a and b such that $a > 0$ and $b > 0$.

Let, $a < b$ also, $a < b$

$a^2 < ab$ (multiplying by $a > 0$ $\Rightarrow ab < b^2$ (multiplying by $b > 0$)

$\Rightarrow a < \sqrt{ab}$ $\Rightarrow \sqrt{ab} < b$

($\because$ $a > 0$, negative sign rejected)

$\Rightarrow a < b$

$\Rightarrow a < \sqrt{ab} < b.$

Thus, the geometric mean of two real positive numbers lies between them. Hence, Fisher's ideal index number which is the geometric mean of Laspeyre's and Paasche's index number, lies between them. More precisely,

If $P_{01}^{La} < P_{01}^{Pa}$ then $P_{01}^{La} < P_{01}^{F} < P_{01}^{La}$

If $P_{01}^{Pa} < P_{01}^{La}$ then $P_{01}^{Pa} < P_{01}^{F} < P_{01}^{La}$

In particular if $P_{01}^{La} = P_{01}^{Pa}$, then all the three indices are equal.

Example 2 :

Construct index numbers of prices from the following data by applying,

1. Laspeyre's method
2. Paasche's method
3. Fisher's Ideal method

	2000		2010	
Commodity	Price	Quantity	Price	Quantity
A	2	8	4	6
B	5	10	6	5
C	4	14	5	10
D	2	19	2	13

Solution :

Commodity	2000		2010		p_1q_0	p_0q_0	p_1q_1	p_0q_1
	Price p_0	Quant. q_0	Price p_1	Quant. q_1				
A	2	8	4	6	32	16	24	12
B	5	10	6	5	60	50	30	25
C	4	14	5	10	70	56	50	40
D	2	19	2	13	38	38	26	26
					$\Sigma p_1q_0 = 200$	$\Sigma p_0q_0 = 160$	$\Sigma p_1q_1 = 130$	$\Sigma p_0q_1 = 103$

1. Laspeyre's Method

$$P_{01} = \frac{\Sigma p_1q_0}{\Sigma p_0q_0} \times 100 = \frac{200}{160} \times 100 = 125$$

2. Paasche's Method

$$P_{01} = \frac{\Sigma p_1q_1}{\Sigma p_0q_1} \times 100 = \frac{130}{103} \times 100 = 126.21$$

3. Fisher's Ideal Index

$$P_{01} = \sqrt{\frac{\Sigma p_1q_0}{\Sigma p_0q_0} \times \frac{\Sigma p_1q_1}{\Sigma p_0q_1}} \times 100 = \sqrt{\frac{200}{160} \times \frac{130}{103}} \times 100 = 125.6$$

CHECK YOUR PROGRESS

1. Define Index number. What is the relation between Laspeyre's and Paasche's index numbers?

2. What is time Reversal Test? Does Fisher's ideal index satisfy time reversal test? Prove.

3. What is factor Reversal test? Does the indices of Laspeyre's and Paasche's indices satisfy this test? Prove.

4. Distinguish between Laspeyre's and Paasche's index number. Show how Fisher's ideal index number is derived from these two.

5. What is chain base index number?

6. Compute index numbers from the following data by using (i) Laspeyre's (ii) Paasche's and (iii) Fisher's ideal index method.

Commodity	Base Year		Current Year	
	Quantity	Price	Quantity	Price
A	12	10	15	12
B	15	7	20	5
C	24	5	20	9
D	5	16	5	14

4.7 Base Shifting :

Base shifting means the changing of the given base period of a series of index numbers and recasting them into a new series based on some recent new base period. The base period can be satisfied to any convenient subsequent period if a particular index number formula satisfies the circular test. Base shifting requires the recomputation of the entire series of the index numbers with the new base. There are two important reasons for shifting the base. They are–

(1) When the base year is too old or too distant from the current period, to make meningful and valid comparisons some recent base period are needed.

(2) If we want to compare series of index numbers with different base periods, to make quick and valid comparisons both the series must be expressed with a common base period.

For base shifting it actually needs the recomputation of the entire series of the index numbers with the new base assigning appropriate weights to the items included in the indices. It is a very difficult task.

One way of tackling the problem of base shifting in a much simpler way consists in taking the index number of the new base year as 100 and then expressing the given series of index numbers as a percentage of the index number to be adopted as new base.

An index number recast to the new base (i.e. new base index number) is :

New base index no. of a year =

$$\frac{\text{Old index no. of the given year}}{\text{Old Index no. of the new base year}} \times 100$$

$$= \left(\frac{100}{\text{Old Index no. of the new base year}}\right) \times$$

(old index no. of the given year)

Thus the new series of index numbers can be obtained on multiplying the old index numbers by the common factor.

In the case of the simple aggregative index (which satisfies the circular test), the base period can be shifted from period 0 to period 2, by using the following relation :

$$P_{2t} = \frac{P_{0t}}{P_{02}} = \frac{\Sigma P_{ti}}{\Sigma P_{0i}} \div \frac{\Sigma P_{2i}}{\Sigma P_{0i}} = \frac{\Sigma P_{ti}}{\Sigma P_{2i}}$$

Such a procedure is not valid in the case of index numbers which donot satisfy the circular test. In particular, however, this limitation is usually ignored and the above procedure is followed viz Expressing the given series of index numbers as a percentage of the index number of the time period selected as the new base year. For example, suppose we have the following series with 1996 as base.

Year	Index Number (Base 1998 = 100)	Index Number
1996	100	$\frac{100}{150} \times 100$
1997	130	$\frac{130}{150} \times 100$
1998	150	$\frac{150}{150} \times 100$
1999	175	$\frac{175}{150} \times 100$
2000	180	$\frac{180}{150} \times 100$

Now, if we wish to shift the base to 1998 it can be done as given in the third column.

4.8 Splicing of Index Number :

Splicing is an application of the principle of base shifting. It consists in combining two or more overlapping series of index numbers to obtain a single continuous series. This continuty of the series of index number is required to facilitate comparisons. In order to secure continuty in comparisons the two series are put together or spliced together to get a continuous series with base shifting, the splicing technique will give accurate results only for the index number which satisfy circular test.

The process of splicing is very simple and is akin to that used in shifting the base.

There may be forward splicing and backward splicing. If an old series is connected with the new one in the sense that the indices of the old series are converted to the base of the new series, it is known as forward splicing. The formula for forward splicing is–

$$\text{Forward spliced Index No.} = \text{Index no. to be spliced} \times \frac{\text{New index no. of the new base year}}{\text{Old index no. of the new base year}}$$

Again, if the new series is connected with the old one in the sense that the indices of the new series are converted to the base of the old series then it is called backward splicing. The formula for backward splicing is,

$$\text{Backward spliced Index No.} = \text{Index no. to be spliced} \times \frac{\text{Old index no. of the new base year}}{\text{New index no. of the new base year}}$$

Example :

We have the following two overlapping series of index number series. Here, either the new series can be spliced to the old one or the old one can be spliced to the new one. This is achieved by the method of proportions.

Year	All India consumer priceIndex (1949 = 100) old series	All India consumer price Index (1954 = 100) new series	New Series spliced to the old one, (Backward)	Old series spliced to the new one (Forward)
1949	100		100	$100 \times \frac{100}{101}$
1950	101		101	$101 \times \frac{100}{101}$
1951	105		105	$105 \times \frac{100}{101}$

1952	103		103	$103 \times \frac{100}{101}$
1953	106		106	$106 \times \frac{100}{101}$
1954	101	100	101	100
1955		97	$97 \times \frac{101}{100}$	97
1956		108	$108 \times \frac{101}{100}$	108
1957		115	$115 \times \frac{101}{100}$	115
1958		120	$120 \times \frac{101}{100}$	120
1959		124	$124 \times \frac{101}{100}$	124

SELF-ASKING QUESTION

Do you think there is any difference between base shifting and splicing of index number? Justify your answer.

4.9 Deflating of Index Numbers :

Deflating means adjusting, correcting or reducing a value which is inflated. Hence by deflating of the price index numbers we mean adjusting them after making allowance for the effect of changing price levels. This technique is extensively used to deflate value series or value indices, rupee sales, income, inventories, wages and so on.

Thus, in a period of rising prices, the increase in the prices of commodities means a fall in the real income of the consumers. Thus, the nominal wages would have to be adjusted for price increases to arrive at the real wages or the deflated income. Thus,

$$\text{Real Wages} = \frac{\text{Money or Nominal Wages}}{\text{Price Index}} \times 100$$

This technique can be used to deflate index number series of sales, inventories etc.

Example :

The consumer price index over a certain period increased from 12(to 215 and the wages of a worker increased from Rs. 1680 to Rs. 3000 What is the gain or loss of the workers?

Solution :

We are given,

Consumer price index in period 1 = 120

Consumer price index in period 2 = 215

The wages of the workers in period 1 and 2 and given to be Rs 1680 and Rs. 3000 respectively.

The real wages of the workers in the current period 2 with respec to the period 1 as base are given by,

$$\frac{120}{215} \times 3000 = 1{,}674.42 \text{ Rs.}$$

Since this wage is less than wages of the workers in period 1 (Rs 1,680) the workers are not better off but worse off by Rs. 5.58 a compared to the period 1.

4.10 Indices of Industrial Production :

The index numbers of industrial production is designed to measur increase or decrease in the level of industrial production in a given perio compared to some base period. Such an index number changes in th quantum of production and not in values. Hence, indices of industria production measures movements in the quantum of production of th individual firms and industries which contribute to the national aggregate Such indices are generally limited to production taking place in secondar industries.

The quantum of production of a firm may be defined as the quantit component of the value added by the firm. Thus, if P and Q denotes th price and quantity of the output and p and q the price and quantity of th input, then the values added by a firm producing a single output with th help of a single input = PQ – pq.

The movement in the value added as between periods 0 and 1 ca be expressed by the ratio.

$$\frac{P_1Q_1 - p_1q_1}{P_0Q_0 - p_0q_0}$$

This ratio can be expressed as the product of its price componen and quantity component as follows :

$$\frac{P_1Q_1 - p_1q_1}{P_0Q_0 - p_0q_0} = \frac{P_0Q_1 - p_0q_1}{P_0Q_0 - p_0q_0} \times \frac{P_1Q_1 - p_1q_1}{P_0Q_1 - p_0q_1} \quad ......(i)$$

$$\text{and} \quad \frac{P_1Q_1 - p_1q_1}{P_0Q_0 - p_0q_0} = \frac{P_1Q_1 - p_1q_1}{P_1Q_0 - p_1q_0} \times \frac{P_1Q_0 - p_1q_0}{P_0Q_0 - p_0q_0} \quad .....(ii)$$

The first term on the right side of both equations (i) and (ii) can be regarded as the quantity component while the second term can be regarded as the price component of the movement in the value added by the firm.

Thus, even with a single input-single output firm, we have two indices of industrial production measuring movement in the quantum of production viz

$$\frac{P_0Q_1 - p_0q_1}{P_0Q_0 - p_0q_0} \text{ and } \frac{P_1Q_1 - p_1q_1}{P_1Q_0 - p_1q_0}$$

The above two measures are the Laspeyre's and the Paasche's types respectively.

In the case of a multi-product firm utilising more than one input, we arrive at the following two measures,

$$N_{01}^{La} = \frac{\Sigma P_0Q_1 - \Sigma p_0q_1}{\Sigma P_0Q_0 - \Sigma p_0q_0}$$

$$N_{01}^{Pa} = \frac{\Sigma P_1Q_1 - \Sigma p_1q_1}{\Sigma P_1Q_0 - \Sigma p_1q_0}$$

Using the same principle the indices of industrial production for individual industries can be combined into an index of industrial production as a whole.

CHECK YOUR PROGRESS

1. What do you mean by deflating using Index numbers ?
   [GU(MA/MSc)Prev '06]
2. Explain splicing and deflating of index number.
   [GU(MA/MSc)Prev '09]
3. What do you mean by indices of industrial production ? Write a note on the indices of Industrial Production.
4. Explain with an example how the base year of index numbers are shifted.

4.11 Summing Up :

An index number is a measure designed to show average changes in the price, quantity or value of a group of items with respect to time, geographical location or situation. In order that the index is representative of the average changes in the level of phenomenon for group, proper weights should be assigned to different commodities according to their respective importance in the group.

According to Irving Fisher, any good index number should satisfy two sets of test viz., the time reversal test and factor reversal test. Another test to check the adequacy of index number is the circular test.

Base shifting means the changing of the given base period of a series of index number and recasting them into a new series based on some new base period.

Splicing is an application of the principle of base shifting. It consists in combining two or more overlapping series of index numbers to obtain a single continuous series.

Deflating of price index numbers means adjusting them after making allowance for the effect of changing price level.

Index numbers of industrial production is designed to measure increase or decrease in the level of industrial production in a given period compared to some base period. They measures movements in the quantum of production of the individual firms and industries which contribute to the national aggregate.

4.12 References and Suggested Readings :

1. Gupta, S.C. and Kapoor, U.K., "Fundamentals of Mathematical Statistics."
2. Gupta, S.C. and Kapoor, V.K., "Fundamentals of Applied Statistics."
3. Gupta, S.C. "Fundamentals of Statistics."
4. Hazarika, P, "Essential Statistics for Economics and Commerce."

●●●