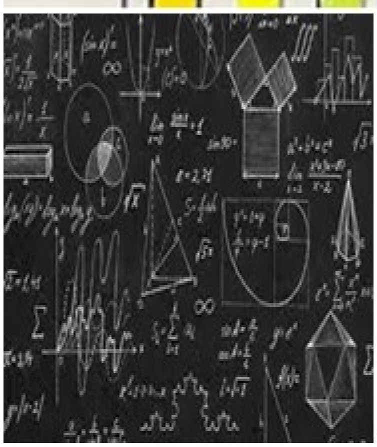
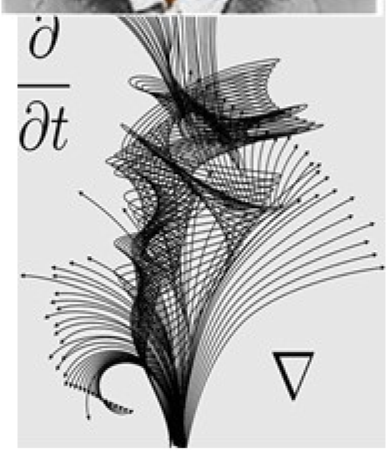
Master of Science
(FIRST SEMESTER)

MAT 503
ADVANCED STATISTICS







DEPARTMENT OF MATHEMATICS
SCHOOL OF SCIENCES
UTTARAKHAND OPEN UNIVERSITY
HALDWANI, UTTARAKHAND
263139

COURSE NAME: ADVANCED STATISTICS

COURSE CODE: MAT 503

Department of Mathematics
School of Science
Uttarakhand Open University
Haldwani, Uttarakhand, India,
263139

BOARD OF STUDIES-FEBRUARY 2023

Editor

Dr. Shivangi Upadhyay
Assistant Professor (AC)
Department of Mathematics,
School of Sciences, Uttarakhand Open University

Unit Writer	Block	Unit
Dr. Arvind Bhatt Associate Professor Department of Mathematics Uttarakhand Open University	**I, II, III, IV**	**01 - 14**

Course Title and Code : ADVANCED STATISTICS (MAT-503)

Edition : 2023

CONTENTS
MAT-503

BLOCK- I

PROBABILITY, RANDOM VARIABLE AND DISTRIBUTIONS

UNIT 1:-BASICS OF PROBABILITY:-

CONTENTS:

1.1.INTRODUCTION:-

Probability is a branch of mathematics that deals with the occurrence of a random event. To study concept of probability, learners should have known the elementary set operations, permutations and combinations. The analysis of events governed by probability is called statistics.

The modern mathematical theory of probability has its roots in attempts to analyze games of chance by Gerolamo Cardano in the sixteenth century, and by Pierre de Fermat and Blaise Pascal in the seventeenth century **(for example the "problem of points").** Christiaan Huygens published a book on the subject in 1657. In the 19th century, what is considered the classical definition of probability was completed by Pierre Laplace.

This culminated in modern probability theory, on foundations laid by Andrey Nikolaevich Kolmogorov. Kolmogorov combined the notion of sample space, introduced by Richard von Mises, and measure theory and presented his axiom system for probability theory in 1933.

23-03-1749 to **05-03-1827**
(Pierre-Simon de Laplace)
https://en.wikipedia.org/wiki/Pierre-imon_Laplace#/media/File:Laplace,_Pierre-Simon,_marquis_de.jpg

Fig 1.1.1

Probability is simply how likely something is to happen. Whenever we're unsure about the outcome of an event, we can talk about the probabilities of certain outcomes—how likely they are. In this unit we are defining Random experiments, Elementary events, Sample Space, Probability, Axiomatic Probability and Translation of events in set theory operation.

- Probability is a feeling of the mind **(Augustus de Morgan)**

Fig 1.1.2
(https://www.cuemath.com/data/terms of-probability/)

1.2 .OBJECTIVES:-

After studying this unit learner will be able to:

1. Describe the notion of probability.
2. Explain the trials and events.
3. Evaluate the probability related to basic random experiments.
4. Construct the example of probability.

1.3.RANDOM EXPERIMENT:-

Consider an experiment whose outcome is not unique, but it has possibility of several outcomes. This type of experiment is known as random experiment. The random experiment can be repeated under identical conditions. Each repetition/ performance is called a trial.

- The example of rolling a dice is a random experiment and each rolling of this dice is a trial.

1.4 .ELEMENTARY EVENT:-

In a trial of a random experiment, any outcome among the all possible outcomes is known as elementary event.

- Tossing a coin give a outcomes as "head" or "tail". Here outcome head is an elementary event, similarly outcome tail is also an elementary event.

1.5. SAMPLE SPACE:-

The collection of all possible outcomes in a random experiments is known as sample space. It is usually denoted as "S".In probability theory, the sample space of an experiment or random trial is the set of all possible outcomes or results of that experiment. A sample space is usually denoted using set notation, and the possible ordered outcomes, or sample points, are listed as elements in the set.

- Rolling a dice and observing the number showing on dice, the sample space is: S={1,2,3,4,5,6}.
- Rolling two dices simultaneously and observing the number showing on dices, the sample space is:S={(1,1), (1,2), (1,3), (1,4), (1,5), (1,6), (2,1), (2,2), (2,3), (2,4), (2,5), (2,6), (3,1), (3,2), (3,3), (3,4), (3,5), (3,6), (4,1), (4,2), (4,3), (4,4), (4,5), (4,6), (5,1), (5,2), (5,3), (5,4), (5,5), (5,6), (6,1), (6,2), (6,3), (6,4), (6,5), (6,6)}.
- Rolling three coins simultaneously and observing the face of coins, the sample space is: S={HHH HHT, HTH, THH, , HTT, THT, TTH, TTT }.
- A sample space can be infinite countable or uncountable set. For example consider a random experiment "tossing a coin until a head comes". Then sample space S = {H,TH,TTH,TTTH,…}.
- Any subset of sample space is known as event. It is usually denoted by capital alphabet A, B, E, E_1, E_2 ,…

- A sample space is called discrete if "S" is countable, otherwise it is called continuous sample space.
- Let the random experiment be "rolling two dices simultaneously". Then the sample space is: S={(1,1), (1,2), (1,3), (1,4), (1,5), (1,6), (2,1), (2,2), (2,3), (2,4), (2,5), (2,6), (3,1), (3,2), (3,3), (3,4), (3,5), (3,6), (4,1), (4,2), (4,3), (4,4), (4,5), (4,6), (5,1), (5,2), (5,3), (5,4), (5,5), (5,6), (6,1), (6,2), (6,3), (6,4), (6,5), (6,6)}. Now consider the following events, E_1 is event that sum of both face is six, E_2 is event that sum of both face is seven, E_3 is event that sum of both face is either six or seven. Then E_1={(1,5), (2,4), (3,3), (4,2), (5,1)}, E_2={(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)} and $E_3 = E_1 \cup E_2$={(1,5),(1,6), (2,4), (2,5), (3,3), (3,4), (4,2), (4,3), (5,1), (5,2), (6,1)}.

  We assign a numerical value to finite discrete sample space and its events as follows:
  $\eta(S) =$ Total number of elements in sample space S,
  $\eta(E) =$ Total number of elements in event E.

CHECK YOUR PROGRESS

Problem 1: Let one card drawn from a pack of 52 cards and E_1be the event that card is red and E_2be the event that card is black ace. Then write total number of elements in $E_1$and E_2.

1.6.PROBABILITY:-

Consider a random experiment. Then it has several events. The numeric value of chance of happening of a particular event in fraction form is known as probability of that event. In case of finite discrete sample space S with equally likely elementary events, the probability of event E is defined as:

$$P(E) = \frac{\eta(E)}{\eta(S)}$$

which is same as:

$$P(E) = \frac{\text{Total number of ways of happening of event E}}{\text{Total number of outcomes}} \quad \ldots.(\mathbf{1.6.1})$$

Total number of ways of happening of event E is number of cases favourable to the event E and also called number of sampling points in E similarly total number of outcomes is number of sample points in E and also called exhaustive number of cases.

- Consider the example 6 given above, then probabilities of events E_1 , E_2 , and E_3 are:$P(E_1) = \frac{\eta(E_1)}{\eta(S)} = \frac{5}{36}$,$P(E_2) = \frac{\eta(E_2)}{\eta(S)} = \frac{6}{36} = \frac{1}{6}$,$P(E_3) = \frac{\eta(E_3)}{\eta(S)} = \frac{11}{36}$.

CHECK YOUR PROGRESS

Problem 2: What is the probability that a leap year contains 53 Monday.
Problem 3: A box contain 1 red, 4 white and 5 black ball. Two ball drawn out of this box simultaneously. What is the probability that one ball is black and one ball is white.

1.7 .TRANSLATION OF EVENTS IN SET THEORY OPERATION:-

Let S be a sample space, $E_1$and $E_2$are events. Then there is formulation of new events usingE_1andE_2. For this consider the following table:

S.No.	Event	Meaning
1.	E_1^c	Event that E_1 does not occur
2.	$E_1 \cup E_2$	Event that E_1 or E_2 occur
3.	$E_1 \cap E_2$	Event that both E_1 and E_2 occur
4.	$E_1 \cap E_2^c$	Event that E_1 occur but E_2does not occur
5.	$E_1 \Delta E_2 = (E_1 \cap E_2^c) \cup (E_1^c \cap E_2)$	Event that exactly one of the events E_1 or E_2 occurs
6.	$E_1^c \cap E_2^c$	Event that none of the events E_1 or E_2 occur
7	$E_1 \cap E_2 = \emptyset$	Events that both E_1 and E_2 are mutually exclusive.
8	Universal set S	Sample space

CHECK YOUR PROGRESS

Problem 4. Let S be a sample space, E_1, $E_2$and E_3 are three arbitrary events. Find expression for the events noted below, in the context of :

(i) Two and more occur. (ii) None Occurs.

1.8.AXIOMATIC PROBABILITY:-

A purely mathematical definition of probability cannot give us the actual value of $P(E)$, the probability cannot give us the actual value of $P(E)$, the probability of occurrence of the event E.

And this must be considered as a function defined on all events.

Definition 1.8.1.(Probability function): Let S be a sample space and Ω be σ-field of events. Then the set function $P: \Omega \rightarrow [0,1]$is said to be probability function, if it satisfies the following conditions:

(i) $P(E) \geq 0$, for every events E in Ω,
(ii) $P(S) = 1$,
(iii) If $E_i, i = 1,2,3, ...$ are mutually disjoint events $(E_i \cap E_j, i \neq j)$, then

$$P(\bigcup_i E_i) = \sum_i P(E_i).$$

Here (S, Ω, P) is called probability space.

Theorem 1.8.1. Probability of the impossible event is zero, i.e.,$P(\emptyset) = 0$.

Proof. Impossible event contains no sample point and hence the certain event S and the impossible event $\emptyset$ are mutually exclusive. Therefore $S \cup \emptyset = S$ it implies that $P(S \cup \emptyset) = P(S)$. Hence, using Axiom of Additivity, we get $P(S) + P(\emptyset) = P(S)$ it implies that $P(\emptyset) = 0$.

Remark 1.8.1: The means of $P(A) = 0$, does not imply that A is necessarily an empty set. In practice, probability '0' is assigned to the events which are so rare that they happen only once in a lifetime.

Theorem1.8. 2. Let E be an event. Then probability of E^c is:
$P(E^c) = 1 - P(E)$ .... ***(1.8.1)***

Proof: Since $(E \cup E^c) = S$, therefore $P(E \cup E^c) = P(S)$.
This implies that$P(E) + P(E^c) = 1$. And hence,$P(E^c) = 1 - P(E)$.

Theorem 1.8. 3. $P(E_1 \cap E_2^c) = P(E_1) - P(E_1 \cap E_2)$ .... ***(1.8.2)***

Proof: Since $(E_1 \cap E_2^c)$ and $(E_1 \cap E_2)$ are disjoint sets whose union is E_1, therefore $P(E_1 \cap E_2^c) + P(E_1 \cap E_2) = P(E_1)$. And hence,$P(E_1 \cap E_2^c) = P(E_1) - P(E_1 \cap E_2)$.

Theorem1.8.4.(Probability of union of two events):

$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$ .... ***(1.8.3)***

Proof: Since $(E_1 \cup E_2)$ can be written as disjoint union of E_1 and $E_1^c \cap E_2$, therefore $P(E_1 \cup E_2) = P(E_1) + P(E_1^c \cap E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$.

Theorem 1.8.5.(For n events): $P(\bigcup_{i=1}^{n} E_i) \geq \sum_{i=1}^{n} P(E_i) - \sum_{1 \leq i < j \leq n} P(E_i \cap E_j)$**……………………………………………………(1.8.4)**

CHECK YOUR PROGRESS

True or false Questions
Problem 5: If something has probability 1,000%, it is sure to happen. T/F
Problem 6: If something has probability 90%, it can be expected to happen about nine times as often as its opposite. T/F

1.9.SOLVED EXAMPLES:-

Example1.9.1: A random arrangement of the letters E,N,G,I,N,E,E,R,I,N,G is done. Find the probability of vowels appear together.

Solution: In the letters E,N,G,I,N,E,E,R,I,N,G there are 3 E's, 2 I's, 2 G's and 3 N's. Therefore total number of different combination formed is: $\frac{11!}{3! \times 2! \times 2! \times 3!}$. Now total number of different combination formed so that vowels appear together is: $\frac{7! \times 5!}{3! \times 2! \times 2! \times 3!}$. Thus the required probability is: $\frac{7! \times 5!}{11!} = \frac{1}{66}$.

Example 1.9.2: From the numbers 1,2,3, ... ,8 two different digits are chosen randomly without replacement. What is the probability that the sum of the digits will be equal to five.

Solution: Total number of ways of choosing two digits from the numbers 1 to 8 is: 8 × 7=56. Total ways of choosing two digits so that sum of the digits will be equal to five are : (1,4),(2,3),(3,2) and (4,1). Hence total number of ways of choosing such that sum will be five is 4. Thus the required probability is: $\frac{4}{56} = \frac{1}{14}$.

Example 1.9.3: From a pack of 52 cards, seven cards are drawn randomly. Find the probability of four cards will be black and three cards will be red?

Solution: Total number of ways of drawing seven card from52 cards is: $\binom{52}{7}$. Now total number of ways of choosing seven card such that four cards will be black and three cards will be red is $\binom{26}{4} \times \binom{26}{3}$. Thus the required probability is: $\frac{\binom{26}{4}\times\binom{26}{3}}{\binom{52}{7}} = \frac{1625}{5593}$. .

Example 1.9.4: Let(S, Ω, P) is a probability space and $\Phi \in \Omega$ be empty set. Then find $P(\Phi)$.

Solution: We know that Φ and S are trivially disjoint sets. Therefore $1 = P(S) = P(S \cup \Phi) = P(S) + P(\Phi) = 1 + P(\Phi)$ this gives$P(\Phi) = 0$.

Example 1.9.5: Let$S = \{1,2,3, \dots, n, \dots\}$be a sample space and Ωbe power set of S and for $E \in \Omega, P(E) = \sum\limits_{n\in E} \frac{1}{2^n}$. Then show that (S, Ω, P) is a probability space.

Solution: Since sum of positive real number is positive, therefore $P(E) \geq 0$, for every events E in Ω. And $P(S) = \sum\limits_{1\leq n\leq\infty} \frac{1}{2^n} = 1$. As we know that in a convergent series rearrangement does not alter the series sum, therefore third axiom follows.

Example 1.9.6: An integer is chosen at random from two hundred digits. What is the probability that the integer is divisible by 6 or 8.

Solution: The sample space of the random experiment is: $S = \{1,2,3, \dots \dots \dots.., 199, 200\} \Rightarrow n(S) = 200$. The event A: 'integer chosen is divisible by 6 ' has the sample points given by : $A = \{6,12,18, \dots \dots.., 198\} \Rightarrow n(A) = \frac{198}{6} = 33$. Therefore $P(A) = \frac{n(A)}{n(B)} = \frac{33}{200}$. Similarly the event B: 'integer chosen is divisible by 8" has the sample points given by : $A = \{8,16,24, \dots \dots.., 200\} \Rightarrow n(B) = \frac{200}{8} = 25$. Therefore $P(B) = \frac{n(A)}{n(B)} = \frac{25}{200}$. The LCM of 6 and 8 is 24. Hence, a number is divisible by both6 and 8, if it is divisible by 24. Therefore $A \cap B = \{24,48,72, \dots \dots.., 192\} \Rightarrow n(A \cap B) = \frac{192}{24} = 8 \implies$ $P(A \cap B) = \frac{8}{200}$.

Hence, the required probability is: from equation ***(1.8.4)*** $P(A \cup B) = \frac{1}{4}$.

Example 1.9.7:A card is drawn from a pack of 52 cards. Find the probability of getting a king or a red card.

Solution: Let us define the following events :

A: the card drawn is a king, B: the card drawn is heart, C: the card drawn is red card. Then $A, B$and C are not mutually exclusive.

$(A \cap B)$: the card drawn is the king of hearts it implies that $n(A \cap B) = 1$. $(B \cap C) = B$: the card drawn is a heart (since $B \subset C$) it implies $n(B \cap C) = 13. C \cap A$:the card drawn is the king of hearts it implies that $n(A \cap B \cap C) = 1$. Therefore $P(A) = \frac{n(A)}{n(B)} = \frac{4}{52}; P(B) = \frac{13}{52}; P(C) = \frac{26}{52}; \ P(A \cap B) = \frac{1}{52}; P(B \cap C) = \frac{13}{52}; P(C \cap A) = \frac{2}{52}; P(A \cap B \cap C) = \frac{1}{52}$.The required probability of getting a king or heart or a red card is given by:

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - \ P(B \cap C) - \ P(C \cap A) + \ P(A \cap B \cap C)$$
$$= \frac{4}{52} + \frac{13}{52} + \frac{26}{52} - \frac{1}{52} - \frac{13}{52} - \frac{2}{52} + \frac{1}{52} = \frac{28}{52} = \frac{7}{13}.$$

Example 1.9.8: An MBA applies for a job in two firms X and Y. The probability of his being selected in firm X is 0.7 and being rejected at Y is 0.5. The probability of his being selected in firm X is 0.7 and being rejected at Y is 0.5. The probability of at least one of his applications being rejected is 0.6. What is probability that he will be selected that he will be selected in one of the firms?

Solution: Let A and B denote the events that the person is selected in firms X and Y respectively. Then in the usual notations, we are given:$P(A) = 0.7 \ \Rightarrow P(\bar{A}) = 1 - 0.7 = 0.3, P(\bar{B}) = 0.5 \Longrightarrow P(B) = 1 - 0.5 = 0.5$ and $P(\bar{A} \ \cup \bar{B}) = 0.6 = P(\bar{A}) + P(\bar{B}) - P(\bar{A} \cap \ \bar{B})$. The probability that the persons will be selected in one of the two firms X and Y is given by: $P(A \cup B) = 1 - P(\bar{A} \ \cup \bar{B}) = 1 - \{P(\bar{A}) + P(\bar{B}) - P(\bar{A} \ \cup \ \bar{B} \ )\} = 1 - (0.3 + 0.5 - 0.6) = 0.8$.

1.10.SUMMARY:-

In this unit, we have studied the basic terminology used in probability. We have also read about the basic idea of probability with some theorems and examples. We have defined probability function. In this unit first we have defined Random experiments, Elementary events and Sample Space with examples. After that we have described the definition of Probability with examples then Axiomatic Probability defined. In this unit Translation of events in set theory operation also defined. This unit is basic outlook of Probability theory and concepts of this unit will be beneficial for the learners in the upcoming units.

1.11. GLOSSARY:-

(i) Set
(ii) Event
(iii) Trial
(iv) Sample space.
(v) Probability.
(Vi) Mutually exclusive.

1.12.REFERENCE BOOKS:-

1. S. C. Gupta and V. K. Kapoor, (2020), *Fundamentals of mathematical statistics*, Sultan Chand & Sons.
2. Seymour Lipschutz and John J. Schiller, (2017), *Schaum's Outline: Introduction to Probability and Statistics*, McGraw Hill Professional.
3. J. S. Milton and J. C. Arnold , (2003*), Introduction to Probability and Statistics (4th Edition),* Tata McGraw-Hill.
4. https://www.wikipedia.org.

1.13. SUGGESTED READINGS:-

1. A.M. Goon,(1998), *Fundamental of Statistics (7th Edition),* 1998.
2. R.V. Hogg and A.T. Craig, (2002), *Introduction to Mathematical Statistics,* MacMacMillan, 2002.
3. Jim Pitman, (1993), *Probability,* Springer-Verlag.
4. https://archive.nptel.ac.in/courses/111/105/111105090

1.14.TERMINAL QUESTIONS:-

TQ 1: Consider a group of 3 men and 2 women and 4 children. From this group four persons are chosen at random. What is the probability of exactly two of them will be children.

TQ 2: What is the probability of a random arrangement of the letters U,N,I,V,E,R,S,I,T,Y, such that two I's do not appear together.

TQ 3: Let E_1, E_2, ... , E_n are n events. Then show that

$$P(\bigcup_{1\le i\le n} E_i) \le \sum_{1\le i\le n} P(E_i)$$

TQ 4: Let E_1, E_2 are twoevents. Then show that

$$P(E_1 \cap E_2) \geq 1 - P(E_1^c) - P(E_2^c)$$

TQ 5: Let $E_1, E_2, \dots, E_n$ are n events. Then show that

$$P(\bigcap_{1\leq i\leq n} E_i) \geq (\sum_{1\leq i\leq n} P(E_i)) - (n-1).$$

TQ6: Define the Probability in your words?..

TQ7: Application of Probability in daily life?...

TQ8 The probability that a learner passes a Physics test is $\frac{2}{3}$ and the probability that He passes both a Physics test and an English test is $\frac{14}{45}$.The probability that he Passes at least one test is $\frac{4}{5}$what is the probability that he passes the EnglishTest?

TQ9 Three newspapers A, B and C are published in a Haldwani, Uttarakhand. It is Estimated from a survey that of the adult population : 20% read A, 16% read B, 14% read$C$, 8% read both $A$and B, 5% read both A and C, 4% read both B and C, 2%read all three. Find what percentage read at least one of the papers?

1.15.ANSWER:-

Answer of Check your progress Questions:-

CYQ 1: $\eta(E_1) = 26$and $\eta(E_2) = 2$.

CYQ2: Probability that a leap year contains 53 Monday is 2/7.

CYQ3: Probability that one ball is black and one ball is white, is 20/45=4/9.

CYQ4: $(i)\left(E_1 \cap E_2 \cap \overline{\acute{E_3}}\right) \cup (\overline{E_1} \cap E_2 \cap E_3) \cup (E_1 \cap \overline{E_2} \cap E_3)$:

$(ii)\left(\overline{E_1} \cap \overline{E_2} \cap \overline{\acute{E_3}}\right)$ **or** $\overline{E_1 \cup E_2 \cup E_3}$

CYQ5: False.

Nothing can have probability 1000 percent -- probability values are restricted to zero to one, or to zero percent to 100 percent. Note that 1.00 = 100 percent, just as one dollar = 100 cents.

CYQ6: True

If something has probability 90 percent, it can be expected to happen about nine times as often as its opposite. Depends on whether "something" and "its opposite" are the only alternatives -- if they are mutually exclusive and collectively exhaustive, then yes, at least as the expected value of a sequence of random trials.

Answer of Terminal Questions:-

TQ 1: 10/21.

TQ2: 4/5.

TQ8: $\frac{4}{9}$.

TQ9: 0.35.

UNIT 2:- CONDITIONAL PROBABILITY

CONTENTS:

2.1. INTRODUCTION:-

Conditional probability is known as the possibility of an event or outcome happening, based on the existence of a previous event or outcome. It is deliberated by multiplying the probability of the preceding event by the renewed probability of the succeeding, or conditional, event. Since we discussed earlier, the probability $P(A)$ of an event A represents the likelihood that a random experiment will result in an outcome in the set A relative to the sample space S of the random experiment. However, quite often, while evaluating some event probability, earlier we have some information emerging from the experiment.

For example, if we have initial information that the outcome of the random experiment must be in a set B of S, then this information must be used to reassess the likelihood that the outcome will also be in B. This reassess probability is indicate by $P(A/B)$and is read as the conditional probability of the event A, given that the event B has already happened.

Ref: https://towardsdatascience.com/deriving-bayes-theorem-the-easy-way-59f0c73496db

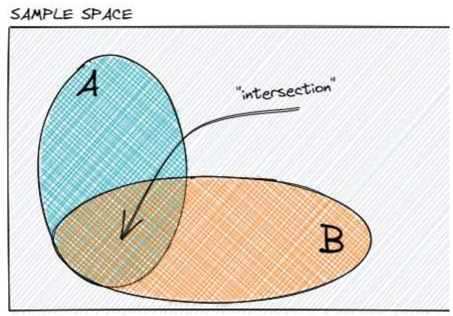


Fig. 2.1.1

Conditional probability clarifies by the following illustration.

- Let us consider a random experiment of drawing a card from a pack of cards. Then the probability of happening of the event A: "The card drawn is a king", is given by: $P(A) = \frac{4}{52} = \frac{1}{13}$.
- Now suppose that a card is drawn and we are informed that the drawn card is red. How does this information affect the likelihood of the event A?. Obviously, if the event B: 'The card is red', has happened, the event 'Black card' is not possible. Hence the probability of the event A must be computed relative to the new sample space 'B' which consists of 26 sample points(red cards only), i.e., $n(B) = 26$. Among these 26 red cards, there are two (red) kings so that $n(A \cap B) = 2$. Hence, the required probability is given by:$P(A/B) = n(A \cap B)/n(B) = 2/26 = 1/13$.

2.2.OBJECTIVES:-

After studying this unit learner will be able to:

1. Explain the notion of Conditional probability.
2. Discuss the independent events.
3. Analyze the concept of Bays' Theorem.
4. Solve the problem related to Conditional probability and Bays' Theorem.

2.3.MULTIPLICATION THEOREM OF PROBABILITY:-

Theorem 2.3.1. For two events A and B,

$$P(A \cap B) = P(A).P(B/A), P(A) > 0$$

$$= P(B).P\left(\frac{A}{B}\right), P(B) > 0 \quad \ldots\ldots\ldots \textbf{(2.3.1)}$$

where $P(B/A)$ represents conditional probability of occurrence of B when the event A has already happened and $P(A/B)$ is the conditional probability of happening of A, given that B has already happened.

Proof. In the usual notation we know that

$P(A) = (\eta(A))/(\eta(S))$, $P(B) = (\eta(B))/(\eta(S))$ and $P(A \cap B) = n(A \cap B)/(n(S))$**......................................(2.3.2)**

For the conditional event A/B, the favourable outcomes must be one of the sample points of B, i.e., for the event A/B, the sample space is B and out of the $n(B)$ sample points, $n(A \cap B)$ pertain to the occurrence of the event A, therefore $P(A/B) = n(A \cap B)/n(B)$. Using equation number (2.3.2)

$P(A \cap B) = n(B)/(n(S)) \times n(A \cap B)/(n(B)) = P(B).P(A/B)$**…...(2.3.3)**

Similarly, we get from (**2.3.2**)

$P(A \cap B) = \frac{n(A)}{n(S)} \times \frac{n(A \cap B)}{n(A)} = P(A).P\left(\frac{B}{A}\right)$ … … … … … … … … … … … … … … … … … … … . .**(2.3.4)**

From ***(2.3.3)*** and ***(2.3.4)*** we get the result ***(2.3.1).***

Therefore "the probability of the simultaneous occurrence of two events A and B is equal to the product of the probability of one of these events and the conditional probability of the other, given that the first one has occurred.

Remark 2.3.1: The conditional probability $P(B/A)$ and $P(A/B)$ are defined if and only if $P(A) \neq 0$ and $P(B) \neq 0$, respectively.

Remark 2.3.2: For $P(B) > 0, P(A/B) \leq P(A)$.

Remark 2.3.3: The conditional probability $P(A/B)$ is not defined if $P(B) = 0$.

Remark 2.3.4: $P(B/B) = 1$.

2.4. INDEPENDENT EVENTS:-

Consider the experiment of throwing two dice, say die 1 and die 2. It is obvious that the occurrence of a certain number of dots on the die 1 has nothing to do with a similar event for the die 2. The two are quite independent of each other, so to say. But suppose, the two dice were connected with a piece of thread before being thrown. The situation changes. This time the two events are not independent in as much as that the uppermost face of one die will have something to do in causing a particular face of the other die to be uppermost; and the shorter the thread, the more is this influence or dependence. Similarly, if we draw two cards from pack of cards in succession, then the results of the two draws are in independent if the cards are drawn with replacement **(i.e., if the first card drawn in placed back in the pack before drawing the second card)** and the results of the two draws are not independent if the cards are drawn without replacement.

2.4.1. DEFINITION OF INDEPENDENT EVENTS AND THEOREMS:-

- Two or more events are said to be independent if the happening or non-happening of any one of them, does not, in any way, affect the happening of others.
- An event A is said to be independent (or statistically independent) of another event B,
  If the conditional probability of Agiven B, i.e., $P(A/B)$ is equal to the unconditional probability of B,
  i.e., if $P(A/B) = P(A)$ ....*(2.4.1)*
  It is important that $P(B) \neq 0$.
- Similarly, If the conditional probability of Agiven B, i.e., $P(B/A)$ is equal to the unconditional probability of B,
  i.e., if $P\left(\frac{B}{A}\right) = P(B)$ … … … … … … ….*(2.4.2)*
  in this case $P(A) \neq 0$.

Theorem 2.4.1. If the events A and B are such that $P(A) \neq 0$, $P(B) \neq 0$ and A is independent of B, then B is independent of A.

Proof. Since the event A is independent of B, therefore $P(A/B) = P(A) \Rightarrow (P(A \cap B))/(P(B)) = P(A) \Rightarrow P(A \cap B) = P(A)P(B)$. Therefore $\frac{P(B \cap A)}{P(A)} = P(B)$[Since $P(A) \neq 0$ and $A \cap B = B \cap A$]. It implies that $P(B/A) = P(B)$ it implies that B is independent of A.

- A is independent of B and B is independent of A it means A and B are independent.
- For any event A in S, A and the null event $\emptyset$ are independent also A and S are independent.
-

Theorem2.4.2.(Multiplication Theorem of Probability for Independent Events).
If the A and B are two events with positive probabilities are such that $P(A) \neq 0$, $P(B) \neq 0$ then$A$ and B are independent if and only if
$P(A \cap B) = P(A)P(B)$………...*(2.4.3)*

Proof. Since, $P(A \cap B) = P(A).P(B/A) = P(B).P(A/B)$; $P(A) \neq 0$,$P(B) \neq 0$…….*(2.4.4)*
A is independent of B and B is independent of A then,
$P(A/B) = P(A)$ and $P(B/A) = P(B)$……………**(2.4.5)**

From ***(2.4.3)*** and ***(2.4.4),*** we get $P(A \cap B) = P(A)P(B)$, as required. Conversely, if ***(2.4.3)*** holds, then we get

$$\frac{P(A \cap B)}{P(B)} = P(A) \Rightarrow P\left(\frac{A}{B}\right) = P(A)$$
$$\frac{P(A \cap B)}{P(B)} = P(B) \Rightarrow P\left(\frac{B}{A}\right) = P(B) \quad \text{.....................................}\textbf{\textit{(2.4.6)}}$$

(2.4.6) implies that A and B are independent events. Hence, for independent events A and B,the probability that both of these occur simultaneously is the product of their respective probabilities. This Rule is known as the Multiplication Rule of Probability.

Theorem 2.4.3. For n events$A_1, A_2, A_3 \dots \dots \dots \dots A_n$ we have
$P(A_1 \cap A_2 \cap \dots \dots \cap A_n) = P(A_1)P(A_2/A_1)P(A_3/(A_1 \cap A_2)) \dots P(A_n/A_1 \cap A_2 \cap \dots \cap \dots \dots A_{n-1})$ ***...........(2.4.7).***
Where $P(A_i/A_j \cap A_k \cap \dots \dots \cap A_l$ represents the conditional probability of the event $A_i, A_{k,} \dots . A_l$ have already happened.

Theorem 2.4.4. Necessary and sufficient condition for independence of$n$ events$A_1, A_2, A_3 \dots \dots \dots \dots A_n$is that the probability of their simultaneous happening is equal to the product of their respective probabilities, i.e.,

$P(A_1 \cap A_2 \cap \dots \dots \cap A_n) = P(A_1)\, P(A_2)P(A_3) \dots P(A_n) \dots \dots$ ***....(2.4.8)***

Theorem 2.4.5. For a fixed B with $P(B) > 0, P(A/B)$is probability fuction.

Theorem 2.4.6. For any three events A, B and C,
$P((A \cup B) / C) = P(A / C) + P(B / C) - P((A \cap B) / C)$.......**(2.4.9)**

Theorem 2.4.7. For any three events A, B and C,
$P((A \cap \overline{B}) / C) + P((A \cap B) / C) = P(A / C)$.......***(2.4.10)***

Theorem 2.4.8. For any three events A, B and C defined on the sample space Ssuch that
$B \subset C$ and $P(A) > 0, P(B/A) \leq P(C/A)$.................***(2.4.11)***

Theorem 2.4.9. If A and B are independent events, then

(i) A and $\overline{B}$ are independent events.
(ii) $\bar{A}$and$B$ are independent events
(iii) $\bar{A}$ and $\bar{B}$ are independent events...............***(2.4.12)***

2.4.2. PAIRWISE INDEPENDENT EVENTS:-

Consider n events $A_1, A_2, A_3 \ldots \ldots \ldots \ldots A_n$ defined on the same sample space so that $P(A_i) > 0; i = 1,2, \ldots \ldots, n$ these events are said to be pair wise independent if every pair of two events is independent in the sense of the definition given in Multiplication Theorem of Probability for Independent Events. The events $A_1, A_2, A_3 \ldots \ldots \ldots \ldots A_n$ are said to be pairwise independent if and only if:

$P(A_i \cap A_j) = P(A_i)P(A_j)$, $i \neq j = 1,2, \ldots \ldots \ldots n. \ldots \ldots \ldots$ ***(2.4.13)***

In particular A_1, A_2, A_3 are said to be pairwise independent if and only if :

$P(A_1 \cap A_2) = P(A_1)P(A_2)$, $P(A_1 \cap A_3) = P(A_1)P(A_3)$,
$P(A_2 \cap A_3) = P(A_2)P(A_3)$.. ***(2.4.14)***

2.4.3. MUTUALLY INDEPENDENT EVENTS:-

Let S denote the sample space for a number of events. The events in S are said to be mutually independent if the probability of the simultaneous occurrence of (any) finite number of them is equal to the product of their separate probabilities.

The events $A_1, A_2, A_3 \ldots \ldots \ldots \ldots A_n$ in a sample space S are said to be mutually independent if:

$$P(A_{i1} \cap A_{i2} \cap \ldots \cap A_{ik}) = P(A_{i1})P(A_{i2}) \ldots P(A_{ik}), \quad k = 2,3, \ldots n \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots . (\mathbf{2.4.15})$$

Hence, the events are mutually independent if they are independent by pairs, and by triplets, and by quadruples, and so on.

Theorem 2.4.10.If A, B, C are mutually independent events then $A \cup B$ and C are also independent.

Theorem 2.4.11 If A, B and C are random events in a sample space and if $, B$ and C are pairwise independent and A is independent of $B \cup C$,then A, B and C are mutually independent.

2.4.4. SOLVED EXAMPLES:-

Example 2.4.4.1. If $A \cap B = \emptyset$, then show that $P(A) \leq P(\bar{B})$.

Solution: Since $A = (A \cap B) \cup (A \cap \bar{B}) = \emptyset \cup (A \cap \bar{B}) = A \cap \bar{B}$
(Since $A \cap B = \emptyset$).

$A \subseteq \bar{B} \Rightarrow P(A) \leq P(\bar{B})$.

* Since $A \cap B = \emptyset$, we have $A \subseteq \bar{B}$, which implies that
$P(A) \leq P(\bar{B})$ … … … … … … … … … … … … …***(2.4.4.1)***

Example 2.4.4.2. Let $A$and B be two events such that $P(A) = \frac{3}{4}$ and $P(A) = \frac{5}{8}$, show that
$(a) P(A \cup B) \geq \frac{3}{4}$,and$(b) \frac{3}{8} \leq P(A \cap B) \leq \frac{5}{8}$.

Solution: We have $A \subset (A \cup B) \Rightarrow P(A \cap B) \leq P(B) = \frac{5}{8}$**……..*(2.4.4.2)***
Also $P(A \cup B) = P(A) + P(B) - P(A \cap B) \leq 1 \Rightarrow \frac{3}{4} + \frac{5}{8} - 1 \leq P(A \cap B)$.Therefore, $\frac{6+5-8}{8} \leq P(A \cap B) \Rightarrow \frac{3}{8} \leq P(A \cap B)$ … … ….***(2.4.4.3)***
From ***(2.4.3.1)*** and ***(2.4.3.2)*** $\frac{3}{8} \leq P(A \cap B) \leq \frac{5}{8}$ ….. .***(2.4.4.4)***

Example 2.4.4.3. For any two events A and B,

$P(A \cap B) \leq P(A) \leq P(A \cup B) \leq P(A) + P(B)$ ….…………..***(2.4.4.5)***

Proof. Since we know that $A = (A \cap \bar{B}) \cup (A \cap B)$. So by the definition of probability
$P(A) = P[(A \cap \bar{B}) \cup (A \cap B)] = P(A \cap \bar{B}) + P(A \cap B)$.Now $P[(A \cap \bar{B})] \geq 0$. Therefore $P(A) \geq P(A \cap B)$ … ….***(2.4.4.6)***
Similarly, $P(B) \geq P(A \cap B)$ it implies that $P(B) - P(A \cap B) \geq 0$.
$P(A \cup B) = P(A) + P(B) - P(A \cap B)$……………..***(2.4.4.7)***
Therefore, $P(A \cup B) \geq P(A)$ it implies that
$P(A) \leq P(A \cap B)$………***(2.4.4.8)***
Also,
$P(A \cup B) \leq P(A) + P(B)$**……………………………………*(2.4.4.9)***
Hence, from ***(2.4.4.6), (2.4.4.7), (2.4.4.8)*** and ***(2.4.4.9)*** we get,
$P(A \cap B) \leq P(A) \leq P(A \cup B) \leq P(A) + P(B)$.

Example 2.4.4.4. The odds against Manager X settling the wage dispute with the workers are $8:6$ and odds in favour of manager Y setting the same dispute are $14:16$.

(i) What is the chance that neither settles the dispute, if they both try, independently of each other?
(ii) What is the probability that dispute will be settled?

Solution:
(i) Let A be the event that the manager X will settle the dispute and B be the event that the manager Y will settle the dispute. Then clearly,
$P(\bar{A}) = \frac{8}{8+6} = \frac{4}{7}$ it implies that $P(A) = 1 - P(\bar{A}) = \frac{6}{14} = \frac{3}{7}$. $P(\bar{B}) = \frac{14}{14+1} = \frac{7}{15}$it implies that $P(A) = 1 - P(B) = \frac{16}{14+1} = \frac{8}{15}$. The required probability that neither settles the dispute is given by:

$P(\bar{A} \cap \bar{B}) = P(\bar{A}) \times P(\bar{B}) = \frac{4}{7} \times \frac{8}{15} = \frac{32}{105}$. (Since A and B are independent it implies that $\bar{A}$and $\bar{B}$ are also independent events).

(ii) The dispute will be settled if at least one of the managers X and Ysettles the dispute. Hence the required probability is given by: $P(A \cup B)$= Probability [at least one of X and Ysettles the dispute.] = 1- Probability [None settles the dispute] = $1 - P(\bar{A} \cap \bar{B}) = 1 - \frac{32}{105} = \frac{73}{105}$.

Example 2.4.4.5.A box contains 6 red, 4 white and 5 black balls. A person draws 4 balls from the box at random. Find the probability that among the balls drawn there is at least one ball of each other.

Solution: The required event E that 'in a draw of 4 balls from the box at random there is at least one ball of each 'color', can materialize in the following mutually disjoint ways:

(i) 1 red, 1 white and 2 black balls; (ii) 2 red, 1 white and 1 black balls; (iii) 1 red, 2 white and 1 black balls. Hence by addition theorem of probability, the required probability is given by:

$$P(E) = P(i) + P(ii) + P(iii)$$

$$= \frac{\binom{6}{1} \times \binom{4}{1} \times \binom{5}{2}}{\binom{15}{4}} + \frac{\binom{6}{2} \times \binom{4}{1} \times \binom{5}{1}}{\binom{15}{4}} + \frac{\binom{6}{1} \times \binom{4}{2} \times \binom{5}{1}}{\binom{15}{4}}$$

$$= \frac{1}{\binom{15}{4}} [6 \times 4 \times 10 + 15 \times 4 \times 5 + 6 \times 6 \times 5]$$

$$= \frac{4!}{15 \times 14 \times 13 \times 12} (240 + 300 + 180) \quad = \frac{24 \times 720}{15 \times 14 \times 13 \times 12}$$

$$= 0.5275.$$

Example 6. Data on readership of a certain magazine show that the proportion of 'male' readers under 35 is 0.40 and over 35 is 0.20. If the proportion of readers under 35 is 0.70, find the proportion of subscribers that are "females over 35 years". Also calculate the probability that a randomly selected male subscriber is under 35 years of age.

Solution. Let us define the following events:
A: Reader of the magazine is a male.
B: Reader of the magazine is over 35 years of age.
Then in usual notations, we are given: $P(A \cap B) = 0.20, P(A \cap \bar{B}) = 0.40andP(\bar{B}) = 0.70 \Rightarrow P(B) = 0.30$.

(i) The proportion of subscribers that are 'female overover35 years' is:

$$P(\bar{A} \cap B) = P(B) - P(A \cap B) = 0.30 - 0.20 - 0.10 = 0.10.$$

(ii) The probability that a randomly selected male subscribers is under 35 years is:

$$P(\bar{B}/A) = \frac{P(A \cap \bar{B})}{P(A)} = \frac{0.40}{0.60} = \frac{2}{3}.$$

[Since, $P(A) = P(A \cap B) + P(A \cap \bar{B}) = 0.20 + 0.40 = 0.60$]

CHECK YOUR PROGRESS

Problem1. An urn contains 4 tickets numbered $1,2,3,4$ and another contains 6 tickets numbered $2, 4, 6, 7, 8, 9$. If one of the two urns is chosen at random and a ticket is drawn at random from the chosen urn, find the probabilities that the ticket drawn bears the number (i) 2 or 4, (ii) 3, (iii) 1 or 9.

Problem 2. From a city population, the probability of selecting (i) a male or a smoker is $\frac{7}{10}$, (ii) a male smoker is $\frac{2}{5}$, and (iii) a male, if a smoker is already selected is $\frac{2}{3}$. Find the probability of selecting (a) a non-smoker, (b) a male (c) a smoker, if a male is first selected.

Problem 3. IfA and B are two events such that $(A \cup B) = \frac{5}{6}, P(A \cap B) = \frac{1}{3}, P(B) = \frac{1}{2}$, then the events A and B are (i) Dependent (ii) Independent (iii) Mutually exclusive (iv) None of these.

Problem 4. For two events $A$and $P(B) = 0.4$, $P(B) = p$, $P(A \cup B) = .6$. Then p equals
(i) 0.2when$A$ and B are mutually disjoint (ii) 0.2 when A and B are independent
(iii) Not determined in any case (iv) 0.2 when A and B are dependent

2.5.BAYES' THEOREM:-

Bayes' theorem which was given by Thomas Bayes, a British Mathematician, in **1763**, provides a means for making these probability calculations .Bayes' Theorem states that the conditional probability of an event, based on the occurrence of another event, is equal to the likelihood of the second event given the first event multiplied by the probability of the first event.

- To prove the Bayes' theorem, use the concept of conditional probability formula.

THOMAS BAYES (1701-1761)
Fig.2.5.1

Ref: https://en.wikipedia.org/wiki/Thomas_Bayes

Theorem 2.5.1. ***Bayes' Theorem.*** If $E_1, E_2, E_3, \dots \dots \dots \dots \dots \dots, E_r$ are mutually disjoint events with $P(E_i) \neq 0, (i = 1,2, \dots \dots \dots \dots, n)$, then for any arbitrary event A which is subset of $\bigcup_{i=1}^{n} E_i$ such that $P(A) > 0$, we have

$P(E_i/A) = \frac{P(E_i)P(A/E_i)}{\sum_{i=1}^{n} P(E_i)P(A/E_i)} = \frac{P(E_i)P(A/E_i)}{P(A)}; i = 1,2, \dots \dots n \dots \dots$***(2.5.1)***

Proof. Since $A \subset \bigcup_{i=1}^{n} E_i$. Since by distributive law we have, $A = A \cap (\bigcup_{i=1}^{n} E_i) = \bigcup_{i=1}^{n}(A \cap E_i)$. Since $(A \cap E_i) \subset E_i, (i = 1,2, \dots \dots, n)$ are mutually disjoint events, we have by addition theorem of probability:

$P(A) = P\{\bigcup_{i=1}^{n}(A \cap E_i)\} = \sum_{i=1}^{n} P\{E_i\}\ P(A/E_i)$**.....................*(2.5.2)***

By multiplication theorem of probability.
Also we have $P(A \cap E_i) = P(A)P(E_i/A)$.
it implies that $P\left\{\frac{E_i}{A}\right\} = \frac{P(A \cap E_i)}{P(A)} = \frac{P(E_i)P(A/E_i)}{\sum_{i=1}^{n} P(E_i)P(A/E_i)}$ [from ***(2.5.2)***].

Remark 2.5.1.

$P(E_1), P(E_2) \dots\dots P(E_n)$	'Prior probabilities.'
$P\left(\frac{A}{E_i}\right), i = 1,2, \dots \dots n$	'likelihoods'
$P\left\{\frac{E_i}{A}\right\}, i = 1,2,3, \dots \dots \dots n$	'Posterior probabilities.'

Remark 2.5.2. If $E_1, E_2, E_3, \dots \dots \dots \dots \dots \dots, E_n$ constitute a disjoint partition of the sample space S and $P(E_i) \neq 0, (i = 1,2, \dots \dots \dots \dots, n)$, then for any arbitrary event A in S we have,

$$P(A) = \sum_{i=1}^{n} P(E_i)P\left(\frac{A}{E_i}\right) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.5.3)$$

Remark 2.5.3.limitations of Bayes Theorem. The Bayesian approach has no general way to represent and handle the uncertainty within the background knowledge and the prior probability function. This is a serious limitation of Bayesianism, both in theory and in application.

Theorem 2.5.4.Bayes' Theorem for Future Events. The probability of the materialisation of another event C,. given $(C/A \cap E_1), P(C/A \cap E_2), \ldots\ldots\ldots\ldots P(C/A \cap E_n)$ is given by:

$$P\left({}^{C}/_{A}\right) = \frac{\sum_{i=1}^{n} P(E_i)P(A/E_i)P(C/E_i \cap A)}{\sum_{i=1}^{n} P(E_i)P(A/E_i)} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.5.4)$$

- Bayes theorem gives the probability of an "event" with the given information on "tests". There is a difference between "events" and "tests".

- For example there is a test for liver disease, which is different from actually having the liver disease, i.e. an event. Rare events might be having a higher false positive rate.

2.6. SOLVED EXAMPLES:-

Example 2.6.1. Suppose that a product is produced in three factoriesX, Y and Z. It is known that factory X produces thrice as many items as factoryY, and that factories Y and Z produce the same number of items. Assume that it is known that 3 per cent of the items produced by each of the factories X and Z are defective while 5 per cent of those manufactured by factory Y are defective. All the items produced in three factories are stoked, and an item of product is selected at random.

(i) What is the probability that this item is defective?
(ii) If an item selected at random is found to be defective, what is the probability that it was produced by factory X, Y and Z respectively?

Solution. Let the number of items produced by each of the factories Y and Z be X, Y and n. Then the number of items produced by the factory X is $3n$. Let E_1, E_2 and E_3 denote the events that the items are produced by factory X, Y and Z respectively and let A be the event of the item being defective. Then we have,
$P(E_1) = \frac{3n}{3n+n+n} = 0.6; P(E_2) = \frac{n}{5n} = 0.2$and $P(E_2) = \frac{n}{5n} = 0.2$.
Also,$P(A/E_1) = P(A/E_3) = 0.03$ and $P(A/E_2) = 0.05$(Given).

(i) The probability that an item selected at random from the stock is defective is given by:

$$P(A) = P\left\{\bigcup_{i=1}^{3}(A \cap E_i)\right\} = \sum_{i=1}^{3} P\{E_i\}\ \mathrm{P}(A/E_i)$$
$$= P(E_1)(A/E_1) + P(E_2)(A/E_2) + P(E_3)(A/E_3)$$
$$= 0.6 \times 0.03 + 0.2 \times 0.05 + 0.2 \times 0.03 = 0.034.$$

(ii) By Bayes' Rule, the required probabilities are given by:

$$\mathrm{P}\,(E_1/A) = \frac{\mathrm{P(E_1)P}(A/E_1)}{\mathrm{P(A)}} = \frac{0.6 \times 0.03}{0.034} = \frac{0.018}{0.034} = \frac{9}{17}.$$
$$\mathrm{P}\,(E_2/A) = \frac{\mathrm{P(E_2)P}(A/E_2)}{\mathrm{P(A)}} = \frac{0.2 \times 0.05}{0.034} = \frac{0.010}{0.034} = \frac{5}{17}.$$
$$\mathrm{P\,(E_3/A)} = \frac{\mathrm{P(E_3)P(A/E_3)}}{\mathrm{P(A)}} = \frac{0.006}{0.034} = \frac{3}{17}.$$

It implies that $\mathrm{P\,(E_3/A)} = 1 - [\mathrm{P}\,(E_1/A) + \mathrm{P}\,(E_2/A)\,] = 1 - \left(\frac{9}{17} + \frac{5}{17}\right) = \frac{3}{17}.$

Example 2.6.2. From a vessel containing 3 white and 5 black balls, 4 balls are transferred into an empty vessel. From this vessel a ball is drawn and is found to be white. What is the probability that out of four balls transferred 3 are white and 1 is black?

Solution. Let us define the following events:

E_1: Transfer of 0 white and 4 black balls.
E_2: Transfer of 1 white and 3 black balls.
E_3: Transfer of 2 white and 2 black balls.
E_4: Transfer of 3 white and 1 black balls
(Since the urn contains 3 white balls, more than 3 white balls cannot be transferred from the vessel)
E: Drawing of a white ball from the second vessel.

Then $P(E_1) = \frac{\binom{5}{4}}{\binom{8}{4}} = \frac{1}{14}$,

$P(E_2) = \frac{\binom{3}{1}\times\binom{5}{3}}{\binom{8}{4}} = \frac{3}{7}$,

$P(E_3) = \frac{\binom{3}{2}\times\binom{5}{3}}{\binom{8}{4}} = \frac{3}{7}$,

$P(E_3) = \frac{\binom{3}{3}\times\binom{5}{1}}{\binom{8}{4}} = \frac{1}{14}$,

Also $\mathrm{P}(E/E_1) = 0,\ \mathrm{P}(E/E_2) = \frac{1}{4}, \mathrm{P}(E/E_3) = \frac{2}{4}$, and $\mathrm{P}(E/E_4) = \frac{3}{4}$.

Hence, by Bayes Theorem, the probability that out of four balls transferred, 3 are white and is black is:

$$P(E_4/E) = \frac{\frac{1}{14}\times\frac{3}{4}}{\frac{1}{14}\times 0+\frac{3}{7}\times 4+\frac{3}{7}\times\frac{1}{2}+\frac{1}{14}\times\frac{3}{4}} = \frac{3}{6+12+3}$$
$$= \frac{1}{7} = 0.14.$$

CHECK YOUR PROGRESS

Problem 5. In 2002 there will be three candidates for the position of principal –Mr.Chatterji, Mr.Ayangar and Dr.Singh-whose chances of getting the appointment are in the proportion $4: 2: 3$ respectively.The probability that Mr.Chatterji if selected would introduce co-education in the college is 0.3 The probabilities of Mr.Ayangar and Dr. Singh doing the same are respectively
0.5 and 0.8.

I. What is probability that there will be co-education in the college in 2003?

II. If there is coeducation in the college in 2003, what is the probability that Dr.Singh is the principal.

2.7. SUMMARY:-

This unit basically based on conditional probability. We are starting with the Introduction and Objectives. In this unit we are giving the definition and proof of Multiplication Theorem of Probability. We also explain the Definition of Independent Events and Theorems, Pairwise Independent Events, Mutually Independent Events. Our main focus in this unit is state and proof of Bayes' Theorem and the problems related with Bayes' Theorem.

2.8.GLOSSARY:-

i. Probability.
ii. Conditional probability.
iii. Independent Events.
iv. Prior probabilities.
v. Likelihoods.
vi. Posterior probabilities.

2.9.REFERENCE BOOKS:-

1. S. C. Gupta and V. K. Kapoor, (2020), *Fundamentals of mathematical statistics*, Sultan Chand & Sons.
2. Seymour Lipschutz and John J. Schiller, (2017), *Schaum's Outline: Introduction to Probability and Statistics*, McGraw Hill Professional.
3. J. S. Milton and J. C. Arnold , (2003*), Introduction to Probability and Statistics (4th Edition),* Tata McGraw-Hill.
4. https://www.wikipedia.org.

2.10.SUGGESTED READINGS

1. A.M. Goon,(1998), *Fundamental of Statistics (7th Edition),* 1998.
2. R.V. Hogg and A.T. Craig, (2002), *Introduction to Mathematical Statistics,* MacMacMillan, 2002.
3. Jim Pitman, (1993), *Probability,* Springer-Verlag.
4. https://archive.nptel.ac.in/courses/111/105/111105090

2.11. TERMINAL QUESTIONS:-

TQ1 State and prove Baye's Theorem?

……………………………………………………………………………
………..

TQ2 What are the criticisms against the use of Baye's Theorem in probability theory?

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

TQ3. One shot is fired from each of the three guns. E_1, E_2, E_3denote the events that the target is hit by the first, second and third guns respectively. If $P(E_1) = 0.5$, $P(E_2) = 0.6$ and $P(E_3) = 0.8$ and E_1, E_2, E_3, are independent events, find the probability that (a) exactly one hit is registered, and (b) at least two hits are registered.

TQ4. Two computers$A$ and B are to be marketed. A salesman who is assigned the job of finding customers for them has 60% and 40% chances respectively of succeeding in case of computer A and B.The two computers can be sold independently. Given that he was able to sell at least one computer, what is the probability that computer A has been sold?

TQ5. The probabilities of X, Y and Z becoming managers are $\frac{4}{9}, \frac{2}{9}$ and $\frac{1}{3}$ respectively. The probabilities that the Bonus Scheme will be introduced if , Y and Z becomes managers are $\frac{3}{10}, \frac{1}{2}$ and $\frac{4}{5}$ respectively

(i)What is the probability that Bonus Scheme will be introduced, and
(ii) if the Bonus Scheme has been introduced, what is the probability that the manager appointed was X?

2.12.ANSWER:-

Answer of Check your progress Questions:-

CYQ 1: $(i)\,\frac{5}{12}\,(ii)\,\frac{1}{8}\,(iii)\,\frac{5}{24}$.

CYQ 2:$(a)\,\frac{3}{5}\,(ii)\,\frac{1}{2}\,(iii)\,\frac{4}{5}$.

CYQ3: Independent.

CYQ 4:.0.2 when$A$ and B are mutually disjoint.

CYQ 5: $(i)\,\frac{23}{45}\,(ii)\,\frac{12}{23}$.

Answer of Terminal Questions:-

TQ 3: $(i)0.26\;\;(ii)0.70$.

TQ4:0.79.

TQ5: $(i)\,\frac{23}{45}\,(ii)\,\frac{6}{23}$.

UNIT 3:- RANDOM VARIABLE

CONTENTS:

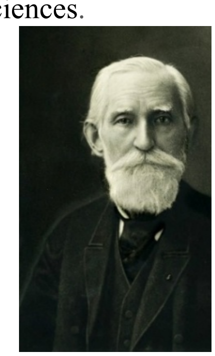
3.1. INTRODUCTION:-

In this unit we are expanding the theory of probability and we are describing the concept random variable. The concept of random variables was introduced by Pafnuty Chebyshev *(1821–1894*), who in the mid-nineteenth century defined a random variable as "a real variable which can assume different values with different probabilities" (Spanos *1999, p. 35).* The concept is closely tied to the theory of probability, which has been studied since the seventeenth century. However, the modern understanding of random variables and their relation to probability arrived more recently, dating to the work by Andrey Kolmogorov *(1933).*Many examples of random variables appear in the social sciences.

Pafnuty Chebyshev (1821–1894),

Ref:https://en.wikipedia.org/wiki/Pafnuty_Chebyshev#/media/File:Pafnuty_Lvovich_Chebyshev.jpg

Fig3.1.1



3.2. OBJECTIVES:-

After studying this unit learner will be able to:

1. Explain the notion of Random variable.
2. Analyse the concept of distribution function.
3. Visualize the Probability Mass/Density Functions

3.3.RANDOM VARIABLE:-

If a real variable X be associated with the outcome of a random experiment, then since the values which X takes depend on chance, it is called a random variable .A rule that assigns a real number to each outcome is called random variable. The rule is nothing but a function of the variable, say, X that assigns a unique value to each outcome of the random experiment. It is clear that there is a value for each outcome, which it takes with certain probability. Thus when a variable X takes the values x_i with probability $p_i(i = 1,2,3,, \dots . n)$, then X is called random variable or a stochastic variable or simply a variate.

- If a random experiment E consists of tossing a pair of dice, the sum X of the two numbers which turn up have the value $2,3,4, \dots \dots \dots . .12$ depending on chance. Then X is the random variable. It is function whose values are real numbers and depend on chance.
- If a random experiment E consists of two tosses the random variable which is the number of heads (0,1or 2).

Outcome	HH	HT	TH	TT
Value of X	2	1	1	0

Thus to each outcome ω, there corresponds a real number $X(\omega)$.Since the points of the sample space S correspond to outcomes, the means that a real number, which we denoted by $X(\omega)$, is defined for each $\omega \in S$.

Definition 3.3.1: Let S be the sample space associated with a given random experiment. A real – valued function defined on S and taking values in $R(-\infty, \infty)$ is called a one – dimensional random variable. If the function values are ordered pairs of real numbers (i.e., vectors in two space), the function is said to be a two – dimensional random variable. More generally, an $n-$ dimensional random variable is simply a function whose domain in S and whose range is a collection of $n-$ tuples of real numbers (vectors in $n-$ space).

Let us consider the probability space, the triplet (S, B, P), where S is the sample space, $viz,$

space of outcomes, B is the $\sigma-$ field of subsets in S, and P is a probability function on B.

Definition 3.3.2:A random variable (r.v) is a function $\mathrm{X}(\omega)$with domain S and range$(-\infty, \infty)$ such that for every real number a, the event $[\omega: \mathrm{X}(\omega) \leq \mathrm{a}] \in \mathrm{B}$.

- One dimensional random variables will be denoted by capital letters, $X, Y, Z, \dots \dots$ etc. A typical outcome of the experiment (i.e., a typical element of the sample space) will be denoted by ω or e. Thus $X(\omega)$ represents the real number which the random variable X associates with the outcome ω. The values which $X, Y, Z, \dots \dots$ etc., can assume are denoted by lower case letters, $viz.$, $x, y, z, \dots \dots \dots$, etc.
- If X_1 and X_2 are random variables and C is a constant then $C_1X_1 + C_2X_2$ is a random variable for constants C_1 and C_2. In particular, $X_1 - X_2$ is a random variable.
- If X is a random variable than
  (i) $\frac{1}{X}$, where $\left(\frac{1}{X}\right)(\omega) = \infty$ if $X(\omega) = 0$,
  (ii)$X_+(\omega) = max\{0, X(\omega)\}$,
  (iii)$X_-(\omega) = max\{0, X(\omega)\}$and
  (iv)$|X|$, are random variable.
- If X_1 and X_2 are random variables, then $(i)max[X_1, X_2]$, and $min[X_1, X_2]$ are also random variables.
- If X is a random variable and $f(.)$ is an increasing function, then $f(X)$ is a random variable.
- If f is a function of bounded variations on every finite interval $[a, b]$, and X is a random variable then $f(X)$ is a random variable.

3.4.DISTRIBUTION FUNCTION:-

Definition3.4.1. Let X be a random variable. The function F defined for all real x by:

$$F(x) = P(X \leq x) = P\{\omega: X(\omega) \leq x\}, -\infty < x < \infty, \dots \dots (\mathbf{3.4.1})$$

is called the distribution function (d.f.) of the random variable (X).

Remark 3.4.1. A distribution function is also called the cumulative distribution function. Sometimes, the notation $F_X(x)$ is used to emphasise the fact that the distribution function is associated with the particular random variable (X). Clearly, the domain of the distribution function is $(-\infty, +\infty)$ and its range is $[0,1]$.

Properties.

- IfF is the distribution function X and if $a < b$, then $P(a < X \leq b) = F(b) - F(a)$.

Proof. The events$'a < X \leq b$' and $'X \leq a'$are disjoint and their union is the event $'X \leq b'$., Hence by addition theorem of probability:

$$\text{P}(a < X \leq b) + P(X \leq a) = P(X \leq b)$$
$$\Rightarrow P(a < X \leq b) = P(X \leq b) - P(X \leq a)$$
$$= F(b) - F(a) \ldots\ldots\ldots\ldots\ldots . (\mathbf{3.4.2})$$

Corollary: $P(a \leq X \leq b) = P\{(X = a) \cup (a < X \leq b)\} = P(X = a) + P(a < X \leq b)\}$

$$= P(X = a) + [F(b) - F(a)] \ldots\ldots\ldots\ldots\ldots .. (\mathbf{3.4.2a})$$

Similarly, we get

$$P(a < X < b) = P(a < X \leq b) - P(X = b)$$
$$= F(b) - F(a) - P(X = b) \ldots\ldots\ldots\ldots . (\mathbf{3.4.2b})$$
$$P(a < X < b) = P(a < X < b) + P(X = a)$$
$$P(a < X < b) = F(b) - F(a) - P(X = b) + P(X = a) \ldots\ldots\ldots\ldots . (\mathbf{3.4.2c})$$

Remark 3.4.2. When $P(X = a) = 0$ and $P(X = b) = 0$, all four events$a \leq X \leq b$,
$a < X < b, a \leq X < b$ and $a < X \leq b$ have the same probability $F(b) - F(a)$.

- If F is distribution function of one dimensional random variable X, then $(i) 0 \leq F(x) \leq 1, (ii) F(x) \leq F(y)$ if $x < y$. In other words, all distribution functions are monotonically non-decreasing and lie between 0&1.
- IfF is distribution function of one dimensional random variable X,then $F(-\infty) = \lim_{X \to -\infty} F(x) = 0$ and $F(\infty) = \lim_{X \to \infty} F(x) = 1$.

Proof. Let us express the whole sample space S as a countable union of disjoint events as follows: $S = \{\cup_{n=1}^{\infty}(-n < X \leq -n + 1)\} \cup \{\cup_{n=0}^{\infty}(n < X \leq n + 1)\}$

$$\Rightarrow P(S) = \sum_{n=1}^{\infty} P(-n < X \leq -n + 1) + \sum_{n=0}^{\infty} P(n < X \leq n + 1)$$
$$\Rightarrow 1 = \lim_{n \to \infty} \sum_{n=1}^{\infty} \{F(-n + 1) - F(-n)\} + \lim_{n \to \infty} \sum_{n=0}^{\infty} \{F(n + 1) - F(n)\}$$
$$= \lim_{n \to \infty} \{F(0) - F(-a)\} + \lim_{n \to \infty} \{F(b + 1) - F(0)\}$$
$$= \{F(0) - F(-\infty)\} + \{F(\infty) - F(0)\}$$

$\therefore 1 = F(\infty) - F(-\infty) \ldots\ldots\ldots\ldots\ldots\ldots\ldots .. (3.4.3)$

$\because -\infty < \infty, F(-\infty) \leq F(\infty)$. Also $F(-\infty) \geq 0$ and Also $F(\infty) \leq 1$.

$\therefore 0 \leq F(-\infty) \leq F(\infty) \leq 1 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (\mathbf{3.4.4})$

From $(\mathbf{3.4.3})$ and $(\mathbf{3.4.4})$, we get $F(-\infty) = 0$ and $F(\infty) = 1$.

Remark 3.4.3. Discontinuities of$F(x)$ are at most countable.

Remark 3.4.4. $F(a) - F(a-0) = \lim_{h\to 0} P(a-h \le X \le a), h < 0$ and $F(a+0) - F(a) = \lim_{h\to 0} P(a \le X \le a+h) = 0, h > 0$.

3.5.DISCRETE RANDOM VARIABLE:-

A variable which can assume only a countable number of real values and for which the value which the variable takes depend on chance, is called a discrete random variable (or discrete stochastic variable or discrete chance variable). A real valued function defined on a discrete sample space is called a discrete random variable.

- Marks obtained in a test, number of accidents per month, number of telephone calls per unit time, number of successes in n trials, and so on.
- A page in a book can have at most 300 words, X= Number of misprints on a page.
  Solution. X = Number of misprints on a page. Since a page in a book has at most 300 words, X takes the finite values. Therefore, random variable X is discrete. Range = $\{0, 1, 2 \ldots .299, 300\}$

3.5.1. PROBABILITY MASS FUNCTION:-

If X is a discrete random variable with distinct values $x_1, x_2, \ldots \ldots \ldots . . x_n, \ldots \ldots \ldots ..$ then the function $p(x)$ defined as:

$$p_X(x) = \begin{cases} P(X = x_i) = p_i, if\ x = x_i \\ 0, if\ x \ne x_i; i = 1,2, \ldots \ldots . \end{cases}$$

is called the probability mass function $(p.m.f.)$ of random variable X. The set of ordered pairs $\{x_i, p(x_i); i = 1,2, \ldots \ldots ., n, \ldots \ldots .\}$ or $\{(x_1, p_1), (x_2, p_2), \ldots \ldots \ldots ., (x_n, p_n); i = 1,2, \ldots \ldots ., n, \ldots \ldots .\}$, specifies the probability distribution of the random variable X.

- The numbers $p(x_i); i = 1,2, \ldots$ must satisfy the following conditions:
  (i) $p(x_i) \ge 0 \forall i, \&(ii) \sum_{i=1}^{\infty} p(x_i) = 1$.
- The set of values which X takes is called the spectrum of the random variable.
  For discrete random variable knowledge of the probability mass function enables us to compute probabilities of arbitrary events. In fact, if E is a set of real numbers. (i) $P(X \in E) = \sum_{x \in E \cap S} p(x)$, where S is the sample space.

3.5.2.DISCRETE DISTRIBUTION FUNCTION:-

The distribution function of a discrete random variable is defined for all real numbers as the probability that the random variable takes a value less than or equal to this real number. In this case there are a countable number of points $x_1, x_2, x_3 \dots \dots$and numer

$$p_i \geq 0, \sum_{i=1}^{\infty} p_i = 1, \text{such that } F(x) = \sum_{i:x_i \leq x} p_i.$$

For example, if $x_i$is just the integer$i$, so that $P(X = x_i) = p_i; i = 1,2,3, \dots \dots \dots$, then $F(x)$
Is a "step function" having jump p, at i, and being constant between each pair of integers.

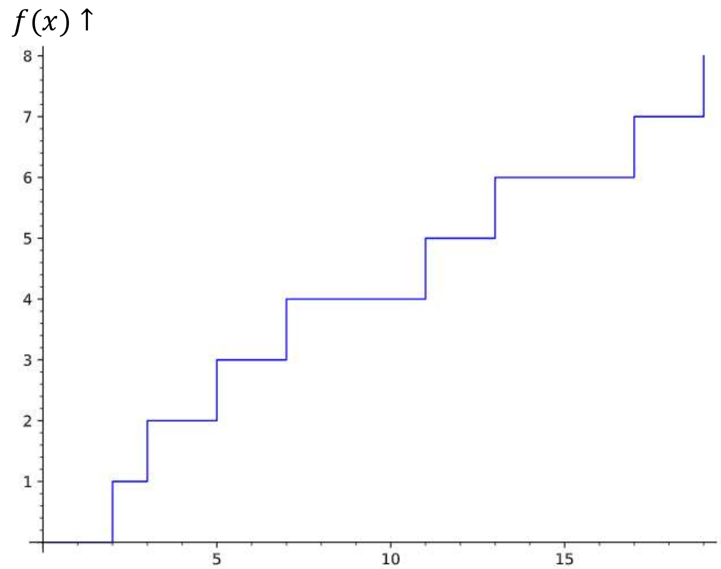


Fig.3.5.2

Theorem 3.5.2.1 $p(x_j) = P(X = x_j) = F(x_j) - F(x_{j-1})$, where F is the distribution function of X.

Proof. Let$x_1 < x_2 < \cdots$We have,
$F(x_j) = P(X \leq x_j) = \sum_{i=1}^{j} p(x_i)$ and $F(x_{j-1}) = P(X \leq x_{j-1}) = \sum_{i=1}^{j-1} p(x_i)$

$$\therefore F(x_j) - F(x_{j-1}) = p(x_j) \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \quad (\mathbf{3.5.2.1})$$

3.6. CONTINUOUS RANDOM VARIABLE:-

A random variable$X$ is said to be continuous if it can take all possible values (integral as well as fractional) between certain limits. In other words, a random variable is said to be continuous when it's different values cannot be put one-one correspondence with a set of positive integers.

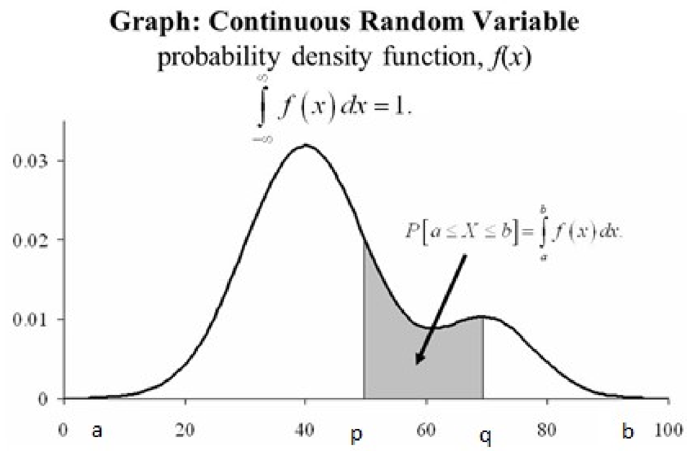


Fig: 3.6.1
https://files.askiitians.com/cdn1/images/201737-12581251-5870-3-random-variables-and-its-probability-distributions.png

The area under the curve y= f(x) bounded by the X-axis and the coordinates x = a and x = b is 1, because it represents the total probability P(a< X <b) which is equal to 1.Also, P (p < X < q) is area under the curve y = f(x) bounded by the X-axis and the coordinates x = p and x = q which is shaded in the figure. The graph of the P. D. F of R. V. X is called the Probability Curve or Probability Density Curve.

Example.age, height, weight, etc.

Example. A gymnast goes to the gymnasium regularly. X = Reduction of his weight in a month.

Solution

X = Reduction of weight in a month, X takes uncountable infinite values,Therefore, random variable X is **continuous**.

3.6.1.PROBABILITY DENSITY FUNCTION:-

Probability Density Function $f_X(x)$of the random variable X is defined as :

$$P(\alpha \le X \le \beta) = \int_{\alpha}^{\beta} f(x)dx \dots\dots\dots\dots.(\mathbf{3.6.1.1})$$

The Probability Density Function$(p.d.f)f_X(x)$ of the random variable X has the following properties:

I. $f(x) \ge 0.$
II. $\int_{-\infty}^{+\infty} f(x)dx = 1.$

III. The probability $P(E)$ given by:

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

In case of continuous random variables the probability at a point is always zero.

$P(x = c) = 0, \forall$ Possible values of c.

$P(\alpha \leq X \leq \beta) = P(\alpha \leq X < \beta) = P(\alpha < X \leq \beta)$

$= P(\alpha < X \leq \beta) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots. (\mathbf{3.6.1.2})$.

3.6.2 .CONTINUOUS DISTRIBUTION FUNCTION:-

If X is continuous random variable with the probability distribution function $f(x)$, then the function:

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt, -\infty < x < \infty \ \ldots\ldots\ldots\ldots. (\mathbf{3.6.2.1})$$

is called the distribution function or sometimes the cumulative distribution of the random variable X.

- $0 \leq F(x) \leq 1, -\infty < x < \infty$.
- $F'(x) = \frac{d}{dx}F(x) = f(x) \geq 0$ [$\because f(x)$ is probability density function].
  $\Longrightarrow F(x)$ is non-decreasing function of x.
- $F(-\infty) = 0, \quad F(\infty) = 1 \Longrightarrow 0 \leq F(x) \leq 1$.
- $F(x)$ is continuous function of x on the right.
- The discontinuities of $F(x)$ are at the most countable.
- $P(a \leq x \leq b) = F(b) - F(a)$.
- $P(a < x < b) = P(a < x \leq b) = P(a \leq x < b) = \int_a^x f(t)dt$.
- $F'(x) = \frac{d}{dx}F(x) \Rightarrow dF(x) = f(x)dx$.
  $dF(x)$ is known as probability differential of X.

3.7.SOLVED EXAMPLES:-

Example 3.7.1. A random variable X has the following probability function:

Values of X, x:	0	1	2	3	4	5	6	7
$p(x)$	0	k	$2k$	$2k$	$3k$	k^2	$2k^2$	$7k^2 + k^2$

i. Find k,
ii. Evaluate $P(X < 6), P(X \geq 6)$, and $P(0 < X < 5)$,

iii. If $P(X \le a) > \frac{1}{2}$, find the minimum value of a, and
iv. Determine the distribution function of X.

Solution.
i. Since $\sum_{x=0}^{7} p(x) = 1$,
$k + 2k + 2k + 3k + k^2 + 2k^2 + 7k^2 + k^2 = 1$.
It implies that $10k^2 + 9k - 1 = 0 \Rightarrow (10k - 1)(k + 1) = 0 \Rightarrow k = \frac{1}{10}$ $or - 1$.
But since $p(x)$ cannot be negative, $k = -1$, is rejected. Hence $k = \frac{1}{10}$.
ii. $P(X < 6) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5)$,

$$= \frac{1}{10} + \frac{2}{10} + \frac{3}{10} + \frac{1}{100} = \frac{81}{100}$$

Now $P(X \ge 6) = 1 - P(X < 6) = 1 - \frac{81}{100} = \frac{19}{100}$
$P(0 < X < 5) = P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) = 8k = \frac{4}{5}$.

X	$F_X(x) = P(X \le x)$
0	0
1	$k = \frac{1}{10}$
2	$3k = \frac{3}{10}$
3	$5k = \frac{5}{10} = \frac{1}{2}$
4	$8k = \frac{4}{5} > \frac{1}{2}$
5	$8k + k^2 = \frac{81}{100}$
6	$8k + 3k^2 = \frac{83}{100}$
7	$9k + 10k^2 = 1$

iii. $(X \le a) > \frac{1}{2}$. By trial, we get $a = 4$.
iv. The distribution function $F_X(x)$ of X is given in the adjoining table.

Example 3.7.2. The diameter of an electric cable , say X, is assume to be a continuous random variable with probability density function :

$$f(x) = 6x(1 - x), 0 \le x \le 1.$$

(i) Check that $f(x)$ is probability density function and
(ii) Determine a number b, such that for $0 \le x \le 1$, $f(x) \ge 0$.

Solution. Obviously, for $0 \le x \le 1, f(x) \ge 0$. Now

$$\int_0^1 f(x)dx = 6\int_0^1 x(1-x)dx = 6\int_0^1 (x-x^2)dx = 6\left|\frac{x^2}{2}-\frac{x^3}{3}\right|_0^1 = 1.$$

Hence $f(x)$ is the probability density function of random variable X.

(ii) $P(X < b) = P(X > b)$

$$\Rightarrow \int_0^b f(x)dx = \int_b^1 f(x)dx \Rightarrow 6\int_0^b x(1-x)dx = 6\int_b^1 x(1-x)dx$$

$$\Rightarrow \left|\frac{x^2}{2}-\frac{x^3}{3}\right|_0^b = \left|\frac{x^2}{2}-\frac{x^3}{3}\right|_b^1 \Rightarrow \left(\frac{b^2}{2}-\frac{b^3}{3}\right) = \left[\left(\frac{1}{2}-\frac{1}{3}\right)-\left(\frac{b^2}{2}-\frac{b^3}{3}\right)\right]$$

$\Rightarrow 3b^2 - 2b^3 = (1 - 3b^2 + 2b^3) \Rightarrow 4b^3 - 6b^2 + 1 = 0 \Rightarrow$

$(2b-1)(2b^2-2b-1) = 0.$

Therefore $2b - 1 = 0 \Longrightarrow b = \frac{1}{2}$ or $2b^2 - 2b - 1 = 0 \Rightarrow b = \frac{2\mp\sqrt{4+8}}{4} = \frac{1\mp\sqrt{3}}{2}$. Hence $b = \frac{1}{2}$, is the only real value lying between 0 and satisfying $\mathbf{3.4.3}$.

Example 3.7.3. A continuous random variable X with probability density function :

$$f(x) = 3x^2, 0 \le x \le 1.$$

(i) Determine a and b, such that (i), $P(X \le a) = P(X > a)$, and (ii), $P(X > b) = 0.05$

Solution. Since a and b, such that

(i), $P(X \le a) = P(X > a)$, each must be equal to $\frac{1}{2}$, because total probability is always unity,

Therefore

$$P(X \le a) = \frac{1}{2} \Rightarrow \int_0^a f(x)dx = \frac{1}{2} \Rightarrow \int_0^a x^2 dx = 3\left|\frac{x^3}{3}\right|_0^a = \frac{1}{2} \Rightarrow a$$

$$= \left(\frac{1}{2}\right)^{\frac{1}{3}}.$$

$$(ii) P(X > b) = 0.05 \Rightarrow \int_b^1 f(x)dx = 0.05 \Rightarrow 3\left|\frac{x^3}{3}\right|_0^1 \Longrightarrow 1 - b^3$$

$$= \frac{1}{20} \Rightarrow b = \left(\frac{19}{20}\right)^{\frac{1}{3}}.$$

Example 3.7.4. A petrol pump is supplied with petrol once a day. If its daily volume of sales (X) in thousands of liters is distributed by

$$f(x) = 5(1-x)^4, 0 \le x \le 1,$$

what must be the capacity of its tank in order that the probability that its supply will be exhausted in a given day shall be 0.01?

Solution. Let the capacity of the tank (in '000 of liters) be $'a'$ such that

$$P(X \ge a) = 0.01 \Rightarrow \int_0^1 f(x)dx = 0.01$$

$\Rightarrow \int_0^1 5(1-x)^4 dx = 0.01$ or $\left[5.\frac{(1-x)^5}{(-5)}\right]_a^1 = 0.01.$

$\Longrightarrow (1-a)^5 = \frac{1}{100}.$

or $1-a = \left(\frac{1}{100}\right)^{1/5} = 1 - 0.3981 = 0.6019.$

Hence the capacity of the tank = $0.6019 \times 1{,}000$ liters = 601.9 liters.

CHECK YOUR PROGRESS

The following statements are True\False:

1. A discrete random variable assigns a whole number to each possible outcome of an experiment. T\F.
2. A continuous random variable cannot assign whole numbers to all the possible outcomes of an experiment. T\F.
3. If $P(X = x) = 0$ for every x, then $X = 0$. T\F.
4. A Continuous random variable is one that can assume an uncountable number of valueT\F
5. To be a legitimate probability density function, all possible values of f(x) must be non-negative. T\F
6. To be a legitimate probability density function, all possible values of f(x) must lie between 0 and 1(inclusive). T\F

3.8. SUMMARY:-

This unit is in important part of Mathematical Statistics. In this unit we are describing the topic random variable and distribution function in briefly. The term random variable is often associated with the idea that value is subject to variations due to chance and a discrete distribution is a random variable defined as a countable variable. In unit we also defined the probability density function and probability mass function.

3.9. GLOSSARY:-

i. Random variable
ii. Distribution function
iii. Probability density function
iv. Probability mass function.

3.10. REFERENCE BOOKS:-

1. S. C. Gupta and V. K. Kapoor, (2020), *Fundamentals of mathematical statistics*, Sultan Chand & Sons.
2. Seymour Lipschutz and John J. Schiller, (2017), *Schaum's Outline: Introduction to Probability and Statistics*, McGraw Hill Professional.
3. J. S. Milton and J. C. Arnold , (2003), *Introduction to Probability and Statistics (4th Edition),* Tata McGraw-Hill.
4. https://www.wikipedia.org.

3.11.SUGGESTED READINGS:-

1. A.M. Goon,(1998), *Fundamental of Statistics (7th Edition),* 1998.
2. R.V. Hogg and A.T. Craig, (2002), *Introduction to Mathematical Statistics,* MacMacMillan, 2002.
3. Jim Pitman, (1993), *Probability,* Springer-Verlag.
4. https://archive.nptel.ac.in/courses/111/105/111105090

3.12.TERMINAL QUESTIONS:-

TQ1. Can you give 5 examples of continuous random variables?
………………………………………………………………………………..............
....

TQ2.What are examples of random variables in everyday life?
………………………………………………………………………………..............
....

TQ3. (i) Is the function defined as follows a density function?

$$f(x) = \begin{cases} e^{-x}, x \geq 0 \\ 0, x < 0. \end{cases}$$

(ii) If so, determine the probability that the variate having this density with fall in the interval $(1,2)$?
(iii) Also find the cumulative probability function?

TQ4. A random variable x has the following probability function:

Values of X, x:	-2	-1	0	1	2	3
$p(x)$	0.1	k	0.2	$2k$	0.3	k

i. Find k.
ii. Calculate mean.
iii. Calculate Variance.

3.13 ANSWER

Answer of Check your progress Questions:-

CYQ1. T

CYQ2. T

CYQ3. F

CYQ4. T

CYQ5. T

CYQ6. F

Answer of Terminal Questions:-

TQ3**.** (i) yes (ii) 0.233 (iii) 0.865.

TQ4. $k = .1; \mu = .8, \sigma^2 = 2.232.$

UNIT 4:- TRANSFORMATIONS AND MOMENT GENERATING FUNCTION

CONTENTS:

4.1 INTRODUCTION:-

In our previous studies we have so far only considered one-dimensional random variables, i.e. we assumed that the outcome of a random experiment could be represented as a single number. However, in many practical situations there are several characteristics associated with the elements of a population or a sample. For example, a physician is interested in several characteristics of a patient, e.g. age, weight, blood pressure, blood sugar values etc. In evaluating the competitiveness of the countries of a certain community, several characteristics are interesting, as e.g. an index of unemployment, stock prices, exchange values etc.In the following we restrict ourselves to the study of only two characteristics, i.e. we consider two-dimensional random variables. In this unit we are explaining about the concept of Two Dimensional Random Variable, Two – Dimensional Distribution Function and notion of transformation of random variable.

4.2 OBJECTIVES:-

After studying this unit learner will be able to:

1. Explain the concept of Two Dimensional Random Variable
2. Evaluate and defined the Two – Dimensional Distribution Function.
3. Analyze the notion of transformation of random variable.

4.3. TWO DIMENSIONAL RANDOM VARIABLE:-

Let Let$X$ and Y be two random variables defined on the sample space S, then the function (X, Y) that assigns in $R^2(= R \times R)$, is called a two-dimensional random variable.If the possible values of (X, Y) are finite or countably infinite, then (X, Y) is called a two-dimensional discrete random variable. When (X, Y) is a two-dimensional discrete random variable the possible values of (X, Y) may be represented as $(x_i, y_j), i = 1,2,3, \dots \dots n, j = 1,2,3, \dots . m$.

Example:4.3.1. Consider the experiment of tossing a coin twice. The sample space is $S = \{HH, HT, TH, TT\}$. Let X denotes the number of heads obtained in the first toss and Y denote the number of heads in the second toss. Then

S	**HH**	**HT**	**TH**	**TT**
X(s)	1	1	0	0
Y(s)	1	0	1	0

$(X, Y)$is a two-dimensional random variable or bi-variate random variable. The range space of $(X, Y)$is $\{(1,1), (1,0), (0,1), (0,0)\}$ which is finite and so $(X, Y)$is a two-dimensional discrete random variables.

4.3.1. JOINT PROBABILITY MASS FUNCTION:-

Let $X$and Ybe two dimensional discrete random variable, defined on the sample space S, then the joint discrete function of X, Y, also called the joint probability mass function of X, Y, denoted by $p_{X,Y}$ is defined as :

$$p_{XY}(x_i, y_j) = \begin{cases} P(X = x_i, Y = y_j) \text{for a value } (x_i, y_j) \text{ of } (X, Y) \\ 0, otherwise \end{cases}$$

……(4.3.1.1)

Remark 4.3.1. It may be noted that $\sum\sum p_{XY}(x_i, y_j) = 1$, where the summation is taken over all possible values of (X, Y).

4.3.2. MARGINAL PROBABILITY FUNCTION:-

Let $X$and Ybe two dimensional discrete random variable, defined on the sample space S, which takes up countable number of values (x_i, y_j). Then the probability distributionof X, is defined asfollows :

$$p_X(x_i) = P(X = x_i)$$

$$=P(X = x_i \cap Y = y_1) + P(X = x_i \cap Y = y_2) + \cdots P(X = x_i \cap Y = y_m)$$

$$= p_{i1} + p_{i2} + p_{i3+\cdots}p_{ij+\cdots}p_{im} = \sum_{j=1}^{m} p_{ij}$$

$$= \sum_{j=1}^{m} p(x_i, y_j) = p_i \ldots\ldots\ldots\ldots (\mathbf{4.3.2.1})$$

and is known as marginal probability mass function or discrete marginal density function X. Also $\sum_{j=1}^{n} p_i = p_1 + p_2 + p_{3+\cdots}p_n = \sum_{j=1}^{n}\sum_{j=1}^{m} p_{ij} = 1$.

Similarly, we can prove that

$p_Y(y_j) = P(Y = y_j) = \sum_{i=1}^{n} p_{ij} = \sum_{i=1}^{n} p(x_i, y_j) = p_j \ldots\ldots\ldots\ldots\ldots\ldots (\mathbf{4.3.2.2})$

which is the marginal probability mass function *of* Y.

4.3.3 CONDITIONAL PROBABILITY FUNCTION:-

Let X and Y be two dimensional discrete random variable, defined on the sample space S. Then the conditional discrete density function or the conditional probability mass function of X, given Y = y, denoted by

$$\mathrm{p_{X/Y}(x/y)} = \frac{\mathrm{P(X = x, Y = y)}}{\mathrm{P(Y = y)}}, \text{provided } \mathrm{P(Y = y)} \neq 0 \ldots\ldots.. (\mathbf{4.3.3.1})$$

Similarly the conditional probability mass function of Y, given$X = x$, denoted by

$$p_{Y/X}(y/x) = \frac{P(X = x, Y = y)}{P(X = X)}, \text{provided } P(X = x) \neq 0 \dots\dots\dots\dots (\mathbf{4.3.3.2})$$

A necessary and sufficient condition for the discrete random variables X and Y to be independent is: $P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j)$ for all values (x_i, y_j) of (X, Y). ……………………………. $(\mathbf{4.3.3.3})$.

4.4. TWO DIMENSIONAL DISTRIBUTION FUNCTION :-

The distribution function of the two dimensional random variable (X, Y) is a real valued function F defined for all real x and y by the relation:

$$F_{XY}(x, y) = P(X \leq x, Y \leq y) \dots\dots\dots\dots (\mathbf{4.3.1})$$

- For the real number a_1, b_1, a_2 and b_2, $P(a_1 < X \leq b_1, a_2 < X \leq b_2)$ $= F_{XY}(b_1, b_2) - F_{XY}(a_1, a_2) - F_{XY}(a_1, b_2) - F_{XY}(b_1, a_2)$.
- For the real number $a_1, b_1, a_2$and b_2, let $a_1 < a_2, b_1 < b_2$, then $(X \leq a_1, Y \leq a_2) + (a_1 < X \leq b_1, Y \leq a_2) = (X \leq b, Y \leq a_2)$ and the events on the $L.H.S$ are mutually exclusive.
- $F(b_1, a_2) - F(a_1, a_2) = P(a_1 < X \leq b_1, Y \leq a_2)$.
- $F(b_1, a_2) \geq F(a_1, a_2)$.
- $F(a_1, b_2) \geq F(a_1, a_2)$.
- $F(x, y)$is monotonic non-decreasing function.
- $F(-\infty, y) = 0 = F(x, -\infty), F(-\infty, +\infty) = 1$.
- If the density function $f(x, y)$ is continuous at $(x, y), \frac{\partial^2 F}{\partial x \partial y} = f(x, y)$.

Two-dimensional distribution (e.g. for two different variables or one variable at two locations or two times). Surface plot indicates multivariate probability density function. Dots indicate sample from the distribution. Both the curves are indicate marginal probability density functions.

4.4.1.MARGINAL DISTRIBUTION FUNCTION :-

For finding the joint distribution function $F_{XY}(x, y)$, it is possible to obtain the individual distribution functions, $F_X(x)$ and $F_Y(y)$ which are termed as marginal distribution fuction of X and Y respectively with respect to the joint distribution function $F_{XY}(x, y)$.

$$F_X(x) = P(X \leq x) = P(X \leq x, Y < \infty)$$
$$= \lim_{y \to \infty} F_{XY}(x, y) = F_{XY}(x, \infty) \ldots \ldots \ldots (\mathbf{4.4.1})$$

Similarly,

$$F_Y(y) = P(Y \leq y) = P(X < \infty, Y \leq y)$$
$$= \lim_{x \to \infty} F_{XY}(x, y) = F_{XY}(\infty, y) \ldots \ldots \ldots (\mathbf{4.4.2})$$

$F_X(x)$ is termed as the marginal distribution function of X corresponding to the joint distribution function $F_{XY}(x, y)$ and similarly $F_Y(y)$ is called marginal distribution function of the random variable Ycorresponding to the joint distribution function $F_{XY}(x, y)$ and similarly $F_Y(y)$ is called marginal distribution function of the random variable Y corresponding to the joint distribution function $F_{XY}(x, y)$. In the case of jointly discrete random variables, the marginal distribution functions are given as:

$F_X(x) = \sum_y P(X \leq x, Y = y)$ and $F_Y(y) = \sum_x P(X = x, Y \leq y)$.Similarly in the case of joint continuous random variable, the marginal distribution functions are given as:

$$F_X(x) = \int_{-\infty}^{x} \left\{ \int_{-\infty}^{\infty} f_{XY}(x, y) dy \right\} dx, F_Y(y)$$
$$= \int_{-\infty}^{y} \left\{ \int_{-\infty}^{\infty} f_{XY}(x, y) dy \right\} dx,$$

4.4.2. JOINT DENSITY FUNCTION, MARGINAL DENSITY FUNCTION :-

From the joint distribution function $F_{XY}(x, y)$ of two-dimensional continuous random variable, the joint probability density function can be evaluated by differentiation as follows:

$$f_{XY}(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}$$
$$= lim_{\delta x \to 0, \delta y \to o} \frac{P(x \leq X \leq x + \delta x, y \leq Y \leq y + \delta y)}{\delta x \delta y} \ldots . (\mathbf{4.4.2.1})$$

"The probability that the point (x, y) will lie in the infinitesimal rectangular region, of area $dxdy$ is given by

$$P\left(x - \frac{1}{2}dx \leq X \leq x + \frac{1}{2}dx, y - \frac{1}{2}dy \leq Y \leq y + \frac{1}{2}dy\right)$$
$$= dF_{XY}(x, y) \ldots . (\mathbf{4.4.2.2})$$

and is denoted by $f_{XY}(x,y)dxdy$, where the function $f_{XY}(x,y)$ is called the joint probability density function of X and Y. The marginal probability function of X and Y are given respectively as follows:

$$f_X(x) = \begin{cases} \sum_y p_{XY}(x,y), (for\ discrete\ variables) \\ \int_{-\infty}^{\infty} f_{XY}(x,y)dy, (for\ continuous\ variables) \end{cases} \dots\dots (\mathbf{4.4.2.3})$$

and $f_Y(y) =$

$$\begin{cases} \sum_x p_{XY}(x,y), (for\ discrete\ variables) \\ \int_{-\infty}^{\infty} f_{XY}(x,y)dx, (for\ continuous\ variables) \end{cases} \dots\dots (\mathbf{4.4.2.4})$$

The marginal probability function of X and Y can be obtained as follows:

$$f_X(x) = \frac{dF_X(x)}{dx} = \int_{-\infty}^{\infty} f_{XY}(x,y)dy, \dots\dots\dots\dots\dots (\mathbf{4.4.2.5})$$

and

$$f_Y(y) = \frac{dF_Y(y)}{dy} = \int_{-\infty}^{\infty} f_{XY}(x,y)dx, \dots\dots\dots\dots\dots\dots (\mathbf{4.4.2.6})$$

Remark 4.4.2: If the joint $p.d.f(p.m.f) f_{XY}(x,y)$ of two random variables X and Y, we can obtain individual distributions of X and Y in the form of their marginal $p.d.f's(p.m.f's)$ $f_X(x)$ and $f_Y(y)$ by using $(\mathbf{4.4.2.3})$and $(\mathbf{4.4.2.4})$. However, the converse is not true it means from the marginal distributions of jointly distributed random variables, we cannot determine the joint distributions of these two random variables.

4.4.3. CONDITIONAL DISTRIBUTION FUNCTION, CONDITIONAL PROBABILITY FUNCTION :-

The Conditional distribution function

$$F_{Y/X}(y/x) = P(Y \le y/X = x) = P(A/X = x) \dots\dots (\mathbf{4.4.3.1})$$

The joint distribution function $F_{XY}(x,y)$ may be f expressed in terms of the conditional distribution as follows.

$$f_{XY}(x,y) = \int_{-\infty}^{x} F_{Y/X}(y/x)dF_X(x) \dots\dots\dots\dots\dots\dots (\mathbf{4.4.3.2})$$

$$f_{XY}(x,y) = \int_{-\infty}^{y} F_{X/Y}(x/y)dF_Y(y) \dots\dots\dots\dots\dots (\mathbf{4.4.3.2})$$

The conditional probability density function of Y given X for two random variables X and Y which are jointly continuous distributed is defined as follows, for two real numbers x and y:

$$f_{Y/X}(y/x) = \frac{\partial}{\partial y} F_{Y/X}(y/x) \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (\mathbf{4.4.3.3})$$

Remark 4.4.3. $f_X(x) > 0$, then $f_{Y/X}(y/x) = \frac{f_{XY}(x,y)}{f_X(x)}$.

Remark 4.4.4. $f_Y(y) > 0$, then $f_{X/Y}(x/y) = \frac{f_{XY}(x,y)}{f_Y(x)}$.

Remark 4.4.5. In terms of the differentials, $P(x < X \leq x + dx/y < Y \leq y + dy) = f_{X/Y}(x/y)dx$.

4.4.4. STOCHASTIC INDEPENDENCE :-

Let us consider two random variables X and Y (of discrete or continuous type) with joint $p.d.f(p.m.f) f_{XY}(x,y)$ and marginal $p.d.f(p.m.f's) f_X(x)$ and $g_Y(y)$ respectively. Then by the compound probability theorem:

$$f_{XY}(x,y) = f_X(x) g_{Y/X}(y/x)$$

where $g_{Y/X}(y/x)$ is the conditional $p.d.f$ of Y for given value of $X = x$.

If we assume that $g_{y/x}$ does not depend on x, then by the definition of marginal $p.d.f(p.m.f's)$, the continuous random variable $g(y) = g(y/x)$.

Two random variables X and Y with joint $p.d.f(p.m.f) f_{XY}(x,y)$ and marginal $p.d.f(p.m.f's) f_X(x)$ and $g_Y(y)$ respectively are said to be stochastically independent if and only if
$$f_{X,Y}(x,y) = f_X(x) g_Y(y) \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (\mathbf{4.4.3.1})$$

- Two jointly distributed random variables X and Y are stochastically independent if and only if their joint distribution function $F_{X,Y}(.,..)$ is the product of their marginal distribution function $F_X(.)$ and $G_Y(.)$ i.e., if for real (x,y)
$$F_{x,y}(x,y) = F_X(x) G_Y(y) \dots\dots\dots\dots\dots (\mathbf{4.4.3.2})$$

- The variables which are not stochastically independent are said to be stochastically dependent.

Theorem 4.4.3. Two random variables X and Y with joint $p.d.f.$ $f(x,y)$ are stochastically independent if and only if $f_{X,Y}(x,y)$ can be expressed as the product of a non-negative function of x alone and a non-negative function of y alone, $i.e.$, if

$$f_{XY}(x,y) = h_X(x) k_Y(y) \dots\dots\dots\dots\dots (\mathbf{4.4.3.3})$$

where $h(.) \geq 0$ and $k(.) \geq 0$.

Proof. If X and Yindependent then by definition $f_{X,Y}(x,y) = f_X(x)g_Y(y)$. Where $f(x)$ and $g(y)$ are marginal $p.d.f's$ of X and Y. Thus condition (4.4.3.3) is satisfied.
Conversely if (4.4.3.3) satisfied, then we prove that X and Y are independent. For continuous random variables X and Y, the marginal $p.d.f's$ are given by:

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy = \int_{-\infty}^{\infty} h(x)k(y)dy = h(x)\int_{-\infty}^{\infty} k(y)dy$$
$$= c_1\, h(x) \ldots\ldots (\mathbf{4.4.3.4})$$

and

$$g_Y(y) = \int_{-\infty}^{\infty} f(x,y)dy = \int_{-\infty}^{\infty} h(x)k(y)dy = k(y)\int_{-\infty}^{\infty} h(x)dy$$
$$= c_2\, k(y) \ldots\ldots (\mathbf{4.4.3.5})$$

wherec_1 and c_2 are constants independent of x and y. Moreover,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dxdy = 1 \Rightarrow \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x)k(y)\,dxdy = 1$$
$$\Rightarrow \left(\int_{-\infty}^{\infty} h(x)dx\right)\left(\int_{-\infty}^{\infty} k(y)dy\right) = 1$$

$\Longrightarrow c_1c_2 = 1$ [From(4.4.3.4) and (4.4.3.5)]……**…(4.4.3.6)**
Therefore we get,

$$f_{XY}(x,y) = h_X(x)k_Y(y) = c_1c_2h_X(x)k_Y(y)$$
$$= \{c_1h_X(x)\}\{c_2k_Y(y)\}$$

From (**4.4.3.4**)and (**4.4.3.5**),

$$f_{XY}(x,y) = f_X(x)g_Y(y)$$

Therefore it implies that X and $Y$are stochastically independent.

If the random variables X and Y are stochastically independent, then for all possible selections of the corresponding pairs of real numbers $(a_1,b_1),(a_2,b_2)$ where $a_i \le b_i$ for all $i = 1,2$ and where the values $\pm\infty$ are allowed, the events $(a_1 < X \le b_1)$and $(a_2 < X \le b_2)$ are independent, i.e.,
$P\{(a_1 < X \le b_1) \cap (a_2 < Y \le b_2)\} = P(a_1 < X \le b_1)P(a_2 < X \le b_2)$….(**4.4.3.7**)

4.5. TRANSFORMATIONS:-

Let (σ,F,P) be a probability space where σ is the set of outcomes, F is collection of events P is the probability measure on the sample space (σ,F). Suppose now that we have a random variable X for the experiment, taking values in a set S, and a function f from S into another set T.Then$Y = f(X)$is a new random variable taking values in T.If the distribution of X is known, how do we find the distribution of Y This is a very basic and important question, and in a superficial sense, the solution is easy. Since for

$B \subseteq T, f^{-1}(B) = \{x \in S: f(x) \in B\}$ is the inverse image of B under $f. P(y \in B) = P[X \in f^{-1}(B)]$for $B \subseteq T$.

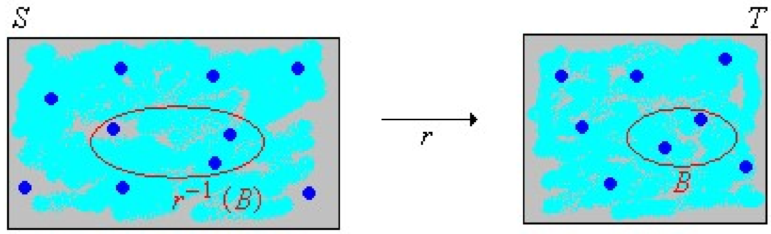


Fig 4.5.1

4.5.1.TRANSFORMATION OF ONE DIMENSIONAL RANDOM VARIABLE:-

Let X be a random variable defined on the event space S and $g(.)$ be a function such that $Y = g(X)$ is also a random variable defined on S.

Theorem 4.5.1. Let X be a continuous random variable with probability density function (p.d.f) $f_X(x)$. Let $y = g(x)$ be strictly monotonic (increasing or decreasing) function of x. Assume that $g(x)$ is differentiable (and continuous) for all x. Then the probability density function (p.d.f) $h(.)$ of the random variable Y is given by :

$$h_Y(y) = f_X(x)\left|\frac{dx}{dy}\right|,$$

where x is expressed in terms of y, and the range of Y isdetermined from the given range of the variable X, on using the transformation $y = g(x)$.
[By the above theorem we shall deal the problem that given the probability density function of a random variable X, to determine the probability density function of a new random variable$Y = g(X)$]

Proof. For solving the proof of this theorem, we are taking the two cases.

Case (i).$y = g(x)$is strictly increasing function of $x\left(i.e, \frac{dy}{dx} > 0\right)$. The distribution function of Y is given by:$H_Y(y) = P(Y \leq y) = P\{g(X) \leq y\} = P\{X \leq g^{-1}(y)\}$, the inverse exists and is unique, since $g(.)$ Is strictly increasing. Therefore $H_Y(y) = F_X\{g^{-1}(y)\}$, where F is the distribution function of $X. H_Y(y) = F_X(x). [\because y = g(x) \Rightarrow g^{-1}(y) = x]$. Differentiating with respect to y, we get

$$h_Y(y) = \frac{d}{dy}\{F_X(x)\} = \frac{d}{dx}\{F_X(x)\}\frac{dx}{dy} = f_X(x)\frac{dx}{dy} \qquad \textbf{.........(4.5.1)}$$

Case (ii).$y = g(x)$is strictly monotonic decreasing function of $x H_Y(y) = P(Y \leq y) = P\{g(X) \leq y\} = P\{X \geq g^{-1}(y)\}$,

$$= 1 - P\{X \geq g^{-1}(y)\} = 1 - F_X\{g^{-1}(y)\} = 1 - F_X(x),$$

where $x = g^{-1}(y)$, the inverse exists and is unique. Differentiating with respect to y, we get,

$h_Y(y) = \frac{d}{dx}\{1 - F_X(x)\}\frac{dx}{dy} = -f_X(x).\frac{dx}{dy} = f_X(x).\frac{-d}{dy}$ ...**(4.5.2).**

It is noted that the algebraic sign (-ve) obtained in **(4.5.2)** is correct, since y is a decreasing function of x it implies that x is a decreasing function of y or $\frac{dx}{dy} < 0$. From equation **(4.5.1)** and **(4.5.2)** $h_Y(y) = f_X(x)\left|\frac{dx}{dy}\right|$.

4.5.2. TRANSFORMATION OF TWO DIMENSIONAL RANDOM VARIABLE:-

Let random variables U and V be transformed to the random variables X and Y by the transformation $u = u(x, y)$, $v = v(x, y)$, where u and v are continuously differentiable functions, for which Jacobian of transformation :

$$J = \frac{\partial(x,y)}{\partial(u,v)} = \begin{vmatrix} \frac{\partial(x)}{\partial(u)} & \frac{\partial(y)}{\partial(v)} \\ \frac{\partial(x)}{\partial(v)} & \frac{\partial(y)}{\partial(v)} \end{vmatrix} \text{.........(4.5.3)}$$

Is either > 0 or < 0 throughout the (x, y) plane so that the inverse transformation is uniquely given by $x = x(u, v)$, $y = y(u, v)$.

Theorem 4.5.2. The joint probability density function (p.d.f.) $g_{UV}(u, v)$ of the transformed variables$U$and V is: $g_{UV}(u, v) = f_{XY}(x, y)|J|$,where $|J|$is the modulus value of the Jacobian of transformation and $f(x, y)$ is expressed in terms of u and v.

Proof. $P(x < X \leq x + dx, y < Y \leq y + dy) = P(u < U \leq u + du, v < V \leq +dv)$.

It implies that $f_{XY}(x, y)dxdy = g_{UV}(u, v)dudv$

$$g_{UV}(u, v)dudv \Rightarrow f_{XY}(x, y)\left|\frac{\partial(x, y)}{\partial(u, v)}\right| = f_{XY}(x, y)dudv$$

$$g_{UV}(u, v) = f_{XY}(x, y)\left|\frac{\partial(x, y)}{\partial(u, v)}\right| = f_{XY}(x, y)|J|.$$

4.6.SOLVED EXAMPLES:-

Example 4.6.1: The joint probability distribution of two random variables X and Y is given by :

$P(X = 0, Y = 1) = \frac{1}{3}$, $P(X = 1, Y = -1) = \frac{1}{3}$, and $P(X = 1, Y = 1) = \frac{1}{3}$.

(i) Find Marginal distribution of X and Y, and

(ii) Conditional probability distribution of X given $Y = 1$.

Solution.

$(\boldsymbol{i})$ $P(X=-1)$

$$=\sum_{y} P(X=-1, Y=y)=P(X=-1, Y=-1)$$
$$+P(X=-1, Y=0)+P(X=-1, Y=1)=0.$$

Similarly $P(X=0)=\frac{1}{3}$

and $P(X=1)=\frac{2}{3}$.

Y \ X	-1	0	1	Marginal (Y)
-1	0	0	$\frac{1}{3}$	$\frac{1}{3}$
0	0	0	0	0
1	0	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{2}{3}$
Marginal (X)	0	$\frac{1}{3}$	$\frac{2}{3}$	1

Thus Marginal distribution of X is:

Values of X, x:	-1	0	1
$P(X=x)$:	0	$\frac{1}{3}$	$\frac{2}{3}$

$(\boldsymbol{ii})$ The conditional probability distribution of X given Y is:

$$P(X=x / Y=y)=\frac{\mathrm{P}(\mathrm{X}=\mathrm{x}, \mathrm{Y}=\mathrm{y})}{\mathrm{P}(\mathrm{Y}=\mathrm{y})}$$

$$P(X=-1 / Y=1)=\frac{\mathrm{P}(\mathrm{X}=-1, \mathrm{Y}=1)}{\mathrm{P}(\mathrm{Y}=1)}=0$$

$$P(X=0 / Y=1)=\frac{\mathrm{P}(\mathrm{X}=0, \mathrm{Y}=1)}{\mathrm{P}(\mathrm{Y}=1)}=\frac{1/3}{2/3}=\frac{1}{2}$$

$$P(X= / Y=1)=\frac{\mathrm{P}(\mathrm{X}=1, \mathrm{Y}=1)}{\mathrm{P}(\mathrm{Y}=1)}=\frac{1/3}{2/3}=\frac{1}{2}.$$

Thus the conditional distribution of X given $Y=1$ is :

Values of X, x:	-1	0	1
$P(X=x / Y=1)$:	0	$\frac{1}{2}$	$\frac{1}{2}$

Example 4.6.2: For the adjoining bivariate probability distribution of two random variables X and Y is find:

i. $P(X \leq 1, Y=2)$,
ii. $P(X \leq 1)$,
iii. $P(Y \leq 3)$and
iv. $(X<3, Y \leq 4)$

Y \ X	1	2	3	4	5	6
0	0	0	$\frac{1}{32}$	$\frac{2}{32}$	$\frac{2}{32}$	$\frac{3}{32}$
1	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$
2	$\frac{1}{32}$	$\frac{1}{32}$	$\frac{1}{64}$	$\frac{1}{64}$	0	$\frac{2}{64}$

Solution: The marginal distributions are given below:

Y \ X	1	2	3	4	5	6	$p_X(x)$
0	0	0	$\frac{1}{32}$	$\frac{2}{32}$	$\frac{2}{32}$	$\frac{3}{32}$	$\frac{8}{32}$
1	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{10}{16}$
2	$\frac{1}{32}$	$\frac{1}{32}$	$\frac{1}{64}$	$\frac{1}{64}$	0	$\frac{2}{64}$	$\frac{8}{64}$
$p_Y(y)$	$\frac{3}{32}$	$\frac{3}{32}$	$\frac{11}{64}$	$\frac{13}{64}$	$\frac{6}{32}$	$\frac{16}{64}$	$\sum p(x) = 1$ $\sum p(y) = 1$

i. $P(X \leq 1, Y = 2) = P(X = 0, Y = 2) + P(X = 1, Y = 2) = \frac{1}{16}$

ii. $P(X \leq 1) = P(X = 0) + P(X = 1) = \frac{8}{32} + \frac{10}{16} = \frac{7}{8}$

iii. $P(Y \leq 3) = P(Y = 1) + P(Y = 2) + P(Y = 3) = \frac{3}{32} + \frac{3}{32} + \frac{11}{64} = \frac{23}{64}$

iv. $(X < 3, Y \leq 4) = P(X = 0, Y \leq 4) + P(X = 1, Y \leq 4) + P(X = 2, Y \leq 4)$
$= \left(\frac{1}{32} + \frac{2}{32}\right) + \left(\frac{1}{16} + \frac{1}{16} + \frac{1}{8} + \frac{1}{8}\right) + \left(\frac{1}{32} + \frac{1}{32} + \frac{1}{64} + \frac{1}{64}\right) = \frac{9}{16}.$

Example 4.6.3: Let $f(x, y) = c(x^2 - y^2)e^{-x}, 0 \leq x < \infty, -x \leq y < x$.

a) Find c.

b) Find the marginal densities.

Solution:

a) $\int_0^\infty \int_{-x}^{x} c(x^2 - y^2)e^{-x}\, dydx = \int_0^\infty \int_{-x}^{x} ce^{-x}(x^2 - y^2)\, dydx$

$$= \int_0^\infty ce^{-x}\left(x^2y - \frac{1}{3}y^3\right)_{y=-x}^{y=x} dx$$

$$= \frac{4c}{3}\Gamma(4)$$

$$= \frac{4c}{3}.3! = 1 \text{ when} c = \frac{1}{8}$$

$$\text{So, } c = \frac{1}{8}.$$

b) $f_Y(y) = \begin{cases} \int_y^\infty f_{XY}(x,y)dx, y \geq 0 \\ \int_{-y}^\infty f_{XY}(x,y)dx, y < 0. \end{cases}$

$$\int_y^\infty f_{XY}(x,y)dx = \frac{1}{8}\int_y^\infty (x^2 - y^2)e^{-x}dx$$

$$= \frac{1}{8}\left[\left\{\int_y^\infty x^2 e^{-x}dx\right\} - \left\{\int_y^\infty y^2 e^{-x}dx\right\}\right]$$

$$= \frac{1}{8}\left[\left\{\int_y^\infty x^2 e^{-x}dx\right\} - y^2 e^{-y}\right]$$

Solving the integration $\int_y^\infty f_{XY}(x,y)dx = \frac{1}{4}e^y(1+y)$ when $y \geq 0$ and $\int_{-y}^\infty f_{XY}(x,y)dx = \frac{1}{4}e^y(1-y)$ when $y < 0$.

Example 4.6.4. For the joint probability distribution of two random variables X and Y given below:

X \ Y	1	2	3	4	Total
1	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	$\frac{10}{36}$
2	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{9}{36}$
3	$\frac{5}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{8}{36}$
4	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	$\frac{5}{36}$	$\frac{9}{36}$
Total	$\frac{11}{36}$	$\frac{9}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	1

I. Marginal distribution of X and Y, and
II. Conditional probability distribution of X given $Y = 1$ and that of Y given the value of $X = 2$.

Solution:**(I)** The marginal distributions of X is defined as:

$$P(X = x) = \sum_y P(X = x, Y = y)$$

Therefore $P(X = 1) = \sum_y P(X = 1, Y = y)$

$$= P(X = 1, Y = 1) + P(X = 1, Y = 2) + P(X = 1, Y = 3) + P(X = 1, Y = 4)$$

$$= \frac{4}{36} + \frac{3}{36} + \frac{2}{36} + \frac{1}{36} = \frac{10}{36}.$$

Similarly $P(X = 2) = \sum_y P(X = 2, Y = y) = \frac{9}{36}$; $P(X = 3) = \sum_y P(X = 3, Y = y) = \frac{8}{36}$

and $P(X = 4) = \sum_y P(X = 4, Y = y) = \frac{9}{36}$.

Similarly, we can obtain the marginal distribution of Y.

Marginal Distribution of *X*

Values of X, x	1	2	3	4
$P(X = x)$	$\frac{10}{36}$	$\frac{9}{36}$	$\frac{8}{36}$	$\frac{9}{36}$

Marginal Distribution of *Y*

Values of Y, y	1	2	3	4
$P(X = y)$	$\frac{11}{36}$	$\frac{9}{36}$	$\frac{7}{36}$	$\frac{9}{36}$

(II) The Conditional probability distribution of X given Y is defined as follows:

$P(X = x/Y = y) = \frac{P(X=x, Y=y)}{P(Y=y)}$. Therefore

$$P(X = 1/Y = 1) = \frac{P(X = 1, Y = 1)}{P(Y = 1)} = \frac{4/36}{11/36} = \frac{4}{11}$$
$$P(X = 2/Y = 1) = \frac{P(X = 2, Y = 1)}{P(Y = 1)} = \frac{1/36}{11/36} = \frac{1}{11}$$
$$P(X = 3/Y = 1) = \frac{P(X = 3, Y = 1)}{P(Y = 1)} = \frac{5/36}{11/36} = \frac{5}{11}$$
$$P(X = 4/Y = 1) = \frac{P(X = 4, Y = 1)}{P(Y = 1)} = \frac{1/36}{11/36} = \frac{1}{11}$$

Hence the conditional distribution of X given $Y = 1$ is :

x:	1	2	3	4
$P(X = x/Y = 1)$	$\frac{10}{36}$	$\frac{9}{36}$	$\frac{8}{36}$	$\frac{9}{36}$

conditional distribution of Y given $X = 2$ as given below:

y:	1	2	3	4
$P(Y = y/X = 2)$	$\frac{1}{9}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{2}{9}$

Example 4.6.5. Let X and Y be jointly distributed with $p.d.f$.:

$$f_{XY}(x, y) = \begin{cases} \frac{1}{4}(1 + xy), |x| < 1, |y| < 1 \\ 0, otherwise \end{cases}$$

Show that X and Y are not independent but X^2 and Y^2 are independent.

Solution. $f_X(x) = \int_{-1}^{1} f(x, y)dy = \frac{1}{4}\left|y + \frac{xy^2}{2}\right|_{-1}^{1} = \frac{1}{2}, -1 < x < 1;$

Similarly, $f_Y(y) = \int_{-1}^{1} f(x, y)dy = \frac{1}{2}, -1 < y < 1;$

$$f_{XY}(x, y) \neq f_X(x)g_Y(y)$$

Therefore it implies that X and Y are not independent. However,

$$P(X^2 \leq x) = P(|X| \leq \sqrt{x}) = \int_{-\sqrt{x}}^{\sqrt{x}} f_X(x)dx = \sqrt{x} \,.. (\mathbf{4.8.5})$$

$$P(X^2 \le x \cap Y^2 \le y) = P(|X| \le \sqrt{x} \cap |Y| \le \sqrt{y})$$
$$= \int_{-\sqrt{x}}^{\sqrt{x}}\left[\int_{-\sqrt{y}}^{\sqrt{y}} f(u,v)dv\right] du = \sqrt{x}\sqrt{y} =$$
$P(X^2 \le x).P(Y^2 \le y)$
Hence, X^2 and Y^2 are independent.

Example 4.6.6. If the cumulative distribution function of a continuous random variable X is $F(x)$, find the cumulative distribution function of :

$$(i) Y = X + a,\quad (ii) Y = X - b\ (iii) Y = aX\ (iv) Y = X^3, and\ (v) Y = X^2.$$

What are the corresponding probability density function?

Solution. Let $G(.)$ be the cumulative distribution function of Y.Then
(i)$G(x) = P(Y \le x) = P(X + a \le x) = P(X \le x - a) = F(x - a)$.
(ii)$G(x) = P(Y \le x) = P(X - b \le x) = P(X \le x + b) = F(x + b)$.
(iii)$G(x) = P(aX \le x) = P(X \le (x/a)) = F(x/a)\ if\ a > 0$and
$G(x) = P(X \ge (x/a)) = 1 - P(X \le (x/a)) = 1 - F(x/a)\ if\ a < 0$
(iv)$G(x) = P(Y \le x) = P(X^3 \le x) = P(X \le x^{1/3}) = F(x^{1/3})$.
(v) $G(x) = (X^2 \le x) = (-x^{1/2} \le X \le x^{1/2}) = P(X \le x^{1/2}) - P(X \le -x^{1/2})$

$$= \begin{cases} 0, if\ x < 0 \\ F(\sqrt{x}) - F(-\sqrt{x} - 0), if\ x > 0. \end{cases}$$

Variable	***Distribution function***	$\boldsymbol{p.d.f.}$
X	$F(x)$	$f(x)$
$X - a$	$F(x + a)$	$f(x + a)$
aX	$\begin{cases} F(x/a), a > 0 \\ 1 - F(x/a), a < 0 \end{cases}$	$\begin{cases} \frac{1}{a} f(x/a), a > 0 \\ -\frac{1}{a} f(x/a), a < 0 \end{cases}$
X^2	$\begin{cases} F(\sqrt{x}) - F(-\sqrt{x} - 0) for\ x \\ 1 - F(x/a), a < 0\ otherw \end{cases}$	$\begin{cases} \frac{1}{2(\sqrt{x})}[f\sqrt{x} + f(-\sqrt{x})], for \\ -0, for\ x \le 0 \end{cases}$
X^3	$F\left(x^{\frac{1}{3}}\right)$	$\frac{1}{3} f\left(x^{\frac{1}{3}}\right).\frac{1}{x^{2/3}}$

Example 4.6.7. Let (X,Y) be a two-dimensional no-negative continuous random variable having the joint density:

$$f(x,y) = \begin{cases} 4xye^{-(x^2+y^2)}; x \ge 0, y \ge 0 \\ 0\ elsewhere \end{cases}$$

Prove that the density function of $U = \sqrt{X^2 + Y^2}$ is:

$$h(u) = \begin{cases} 2u^3 e^{-u^2}, 0 \le u < \infty \\ 0, elsewhere \end{cases}$$

Solution. Let us make the transformation $u = \sqrt{x^2 + y^2}$ and $v = x$.

$\Rightarrow v \geq 0, u \geq 0andu \geq v$ or $u \geq 0$ and $0 \leq v \leq u$.
The Jacobian of transformation J is given by:

$$\frac{1}{J} = \frac{\partial(u,v)}{\partial(x,y)} = \begin{vmatrix} \frac{\partial(u)}{\partial(x)} & \frac{\partial(v)}{\partial(x)} \\ \frac{\partial(u)}{\partial(y)} & \frac{\partial(v)}{\partial(y)} \end{vmatrix} = \begin{vmatrix} \frac{x}{\sqrt{x^2+y^2}} & 1 \\ \frac{y}{\sqrt{(x^2+y^2)}} & 0 \end{vmatrix} = \frac{-y}{\sqrt{x^2+y^2}}$$

The joint $p.d.f$ of U and V is given by:$g_{UV}(u,v) = f_{XY}(x,y)|J| = 4xye^{-(x^2+y^2)}\left|-\frac{\sqrt{x^2+y^2}}{y}\right| = 4x\sqrt{x^2+y^2}e^{-(x^2+y^2)}$

$$= \begin{cases} 4vu.e^{-u^2}; u \geq 0, 0 \leq v \leq u \\ 0, \quad otherwise \end{cases}$$

Hence the marginal density function $U = \sqrt{X^2+Y^2}$ is:

$$h(u) = \int_0^u g(u,v)dv = 4u.e^{-u^2}\int_0^u vdv = \begin{cases} 2u^3e^{-u^2}, u \geq 0 \\ 0, elsewhere \end{cases}$$

Check your Progress

Problem1. The joint probability density of the random variable X and Y is:

$f(x,y) = \begin{cases} \frac{1}{2}e^{-|x|-|y|, -\infty<x<\infty} \\ 0, -\infty < y < \infty \end{cases}$, The probability that $x \leq 1$ and $y \leq 0$.

a) $\frac{1}{2}(2-e^{-1})$
b) $\frac{1}{2}(2-e^{-2})$
c) $\frac{1}{4}(2-e^{-1})$
d) None of the above.

Problem2. Two random variables X and Y are distributed according to …

$$f_{X,Y}(x,y) = \begin{cases} (x+y) & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ & 0, otherwise \end{cases}$$

The probability $(X+Y \leq 1)$ is…………

a) 0.44
b) 0.22
c) 0.33
d) None of the above.

Problem3.. The joint pdf of a bivariate random variable (X,Y) is given by

$$f_{X,Y}(x,y) = \begin{cases} kxy & 0 < x < 1, 0 < y < 1 \\ & 0, otherwise \end{cases}$$

Where k is constant. The value of k is……..

a) 1
b) 2
c) 3
d) 4

Problem4. The condition al probability mass function of X, given Y = y, …………

Problem5..The joint probability density function (p.d.f.) $g_{UV}(u,v)$ of the transformed variables U and V is…………………………

4.7. SUMMARY:-

In the previous unit we studied various aspects of the theory of a single random variable. In this unit we extend our theory to include two random variables one for each coordinator axis X and Y of the XY Plane. The section of transformation studies how the distribution of a random variable changes when the variable is transformed in a deterministic way. Basically this unit is complete overview concept of Two Dimensional Random Variable, Two – Dimensional Distribution Function and transformation of random variable.

4.8. GLOSSARY:-

1. Probability
2. Sample space
3. Random Variable
4. Two Dimensional Random Variable
5. Two – Dimensional Distribution Function
6. Probability density function
7. Probability mass function
8. Transformations

4.9. REFERENCE BOOKS:-

1. S. C. Gupta and V. K. Kapoor, (2020), *Fundamentals of mathematical statistics*, Sultan Chand & Sons.
2. Seymour Lipschutz and John J. Schiller, (2017), *Schaum's Outline: Introduction to Probability and Statistics*, McGraw Hill Professional.
3. J. S. Milton and J. C. Arnold , (2003), *Introduction to Probability and Statistics (4th Edition),* Tata McGraw-Hill.
4. https://www.wikipedia.org.

4.10. SUGGESTED READINGS:-

1. A.M. Goon,(1998), *Fundamental of Statistics (7th Edition),* 1998.
2. R.V. Hogg and A.T. Craig, (2002), *Introduction to Mathematical Statistics,* MacMacMillan, 2002.
3. Jim Pitman, (1993), *Probability,* Springer-Verlag.
4. https://archive.nptel.ac.in/courses/111/105/111105090

4.11.TERMINAL QUESTIONS:-

TQ1.A two – dimensional random variable (X, Y) have a bivariate distribution given by: $P(X = x, Y = y) = \frac{x^2+y}{32}$, for $x = 0,1,2,3$ and $y = 0,1$. Find the marginal distribution of X and Y.

TQ2. A two – dimensional random variable (X, Y) have a joint probability mass function: $p(x, y) = \frac{1}{27}(2x + y)$, where x and y can assume only the integers values 0,1 and 2. Find the conditional distribution of Y for $X = x$.

TQ3. Given $f(x, y) = e^{-(x+y)}; 0 \le x < \infty, 0 \le y < \infty$. Are X and Y independent?Find $(i)P(X > 1), (ii)P(X < Y/X < 2Y)(iii)P(1 < X + Y < 2)$.

TQ4. Given the joint density function ofX and Yas:

$$f(x, y) = \begin{cases} \frac{1}{2}xe^{-y}; 0 < x < 2, y > 0 \\ 0 \; elsewhere \end{cases}$$

Find the distribution of X + Y.

4.12.ANSWERS:-

Answer of Check your progress Questions:

CYQ 1: a)$\frac{1}{2}(2 - e^{-1})$.

CYQ2: c) 0.33.

CYQ3: d) 4.

CYQ 4: $p_{X/Y}(x/y) = \frac{P(X=x,Y=y)}{P(Y=y)}$, provided $P(Y = y) \neq 0$.

CYQ 5: $g_{UV}(u, v) = f_{XY}(x, y)|J|$.

Answer of Terminal Questions:-

TQ1. Marginal distribution of $X$and Y.

Marginal distribution of $P(X = x)$	$\frac{1}{32}$	$\frac{3}{32}$	$\frac{9}{32}$	$\frac{19}{32}$	1

Marginal distribution of $P(Y = y)$	$\frac{14}{32}$	$\frac{18}{32}$	1

TQ2.

Y \ X	0	1	2
1	0	$\frac{1}{3}$	$\frac{2}{3}$
2	$\frac{2}{9}$	$\frac{3}{9}$	$\frac{4}{9}$
3	$\frac{4}{15}$	$\frac{5}{15}$	$\frac{6}{15}$

TQ3. $(i)\frac{1}{e}\,(ii)\frac{3}{4}\,(iii)\frac{2}{e}-\frac{3}{e^2}$

TQ4. $g(u) = \begin{cases} \frac{1}{2}(e^{-u}+u-1), 0 < u \le 2 \\ \frac{1}{2}e^{-u}(1+e^2), 2 < u < \infty \\ 0 \text{ elsewhere} \end{cases}$

BLOCK- II

GENERATING FUNCTION AND LAW OF LARGE NUMBERS

UNIT 5:- MATHEMATICAL EXPECTATION AND MOMENT GENERATION FUNCTION

CONTENTS:

5.1.INTRODUCTION:-

In this unit we are explaining Mathematical expectation, Moments, Moment generating function and Characteristic function. The general idea of generating function has much wider scope than its applications to probability. The proper setting is ``harmonic analysis" which is one of the central and most developed parts of mathematics. The birth of the idea can be traced back to Abraham de Moivre ***(1667-1754),*** and his book Doctrine of Chances. Later the same idea was developed and applied in number theory (Euler), and most importantly in mathematical physics (Fourier). (Characteristic function is a special case of Fourier transform). Laplace transform (the moment generating function) belongs to the same circle of ideas, and its original use was also in probability. All this powerful set of ideas (Generating function, Fourier, Laplace and other similar transforms) is called (linear) Harmonic analysis and it is one of the most powerful methods of mathematics, with applications practically everywhere (in almost all areas of mathematics and sciences).

Abraham de Moivre
26 May 1667 – 27 November 1754)
https://en.wikipedia.org/wiki/Abraham_de_Moivre#/media/File:Abraham_de_moivre.jpg

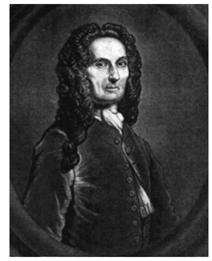


Fig.5.1.1

5.2. OBJECTIVES:-

After studying this unit learner will be able to:

.

1. Explain the Mathematical Expectations and Moments.
2. Understand the Moment Generating Function
3. Defined the characteristic function.
4. Evaluate the moment generating function and characteristic function.
5. Describe the notion of transformation of random variable.

5.3.MOMENTS:-

Moments are a set of statistical parameters to measure a distribution. Moments are statistical tools, used in statistical investigations. The moments of a distribution are the arithmetic means of the various powers of the deviations of items from some given number. In simple terms, the moment is a way to measure how spread out or concentrated the number in a dataset is around the central value, such as the mean.

Moments about an arbitrary number (Raw Moments): The rth moment of a variable x about any point $x = A$, usually denoted by $\mu_r'(Raw\ Moments)$ is given by

$$\mu_r' = \frac{1}{N}\sum_i f_i\,(x_i - A)^r, \sum_i f_i = N \ldots\ldots\ldots\ldots\ldots.(\mathbf{5.3.1})$$

(For an Individual Series)

$$\mu_r' = \frac{1}{N}\sum_i f_i\,(d_i)^r, \text{where} d_i = x_i - A \ldots\ldots\ldots\ldots(5.3.2)$$

(For a Frequency Distribution)

Moments about Mean (Central Moments): The rth moment of a variable x about any point $\bar{x}$, usually denoted by μ_r(Central Moments) is given by:

$$\mu_r = \frac{1}{N}\sum_i f_i\,(x_i - \bar{x})^r, \sum_i f_i = N \ldots\ldots\ldots\ldots\ldots.(\mathbf{5.3.3}))$$

(For an Individual Series)

$\mu_r' = \frac{1}{N}\sum_i f_i\,(z_i)^r$,where$z_i = x_i - \bar{x}$*.........(5.3.4)*
*(*For a Frequency Distribution)
Therefore $\mu_0 = \frac{1}{N}\sum_i f_i\,(x_i - \bar{x})^0 = \mu_r = \frac{1}{N}\sum_i f_i = \frac{1}{N}.N = \mu_0 = 1$
and $\mu_1 = \frac{1}{N}\sum_i f_i\,(x_i - \bar{x}) = \frac{1}{N}\sum_i f_i\,x_i - \frac{1}{N}\sum_i f_i\,\bar{x} = \bar{x} - \frac{1}{N}N.\bar{x} = 0$.$\mu_2 = \frac{1}{N}\sum_i f_i\,(x_i - \bar{x})^2 = \sigma^2$.
If $d_i = x_i - A$, then $\bar{x} = A + \frac{1}{N}\sum_i f_i\,d_i = A + \mu_1'$**..............*(5.3.5)***

Relation Between Moments about Mean in Terms of Moments about Any Point and Vice Versa.

$\mu_r = \frac{1}{N}\sum_i f_i\,(x_i - \bar{x})^r = \mu_r = \frac{1}{N}\sum_i f_i\,(x_i - A + A - \bar{x})^r = \frac{1}{N}\sum_i f_i\,(d_i + A - \bar{x})^r$where$d_i = x_i - A$. Using (4.4.5), we get

$$\mu_r = \frac{1}{N}\sum_i f_i\,(d_i - \mu_1')^r$$

$$= \frac{1}{N}\sum_i f_i\,(d_i^r - \binom{r}{1} d_i^{r-1}\mu_1' + \binom{r}{2} d_i^{r-2}(\mu_1')^2 - \binom{r}{3} d_i^{r-3}(\mu_1')^3 + \cdots (-1)^r(\mu_1')^r$$

$$\mu_r = \mu_r' - \binom{r}{1}\mu_{r-1}'\mu_1' + \binom{r}{1}\mu_{r-2}'\mu_{r-2}' - (-1)^r(\mu_1')^r \quad \textbf{.........}\textit{(5.3.6)}$$

In particular, on putting $r = 2, 3$ & 4 in ***(5.3.6)*** and simplifying, we get
$\mu_2 = \mu_2' - {\mu_1'}^2$,
$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2{\mu_1'}^3$,
$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'{\mu_1'}^2 - 3{\mu_1'}^4$
Simililary,

$$\mu_r' = \frac{1}{N}\sum_i f_i\left(z_i^r + \binom{r}{1}z_i^{r-1}\mu_1' + \binom{r}{2}z_i^{r-2}{\mu_1'}^2 \ldots\ldots\ldots\ldots.{\mu_1'}^r\right)$$
$$= \frac{1}{N}\sum_i f_i\left(z_i^r + \binom{r}{1}z_i^{r-1}\mu_1' + \binom{r}{2}z_i^{r-2}{\mu_1'}^2 \ldots\ldots\ldots{\mu_1'}^r\right)$$
$$= \mu_r + \binom{r}{1}\mu_{r-1}\mu_1' + \binom{r}{2}\mu_{r-2}{\mu_1'}^2 + \cdots{\mu_1'}^r.$$ From ***(5.3.5)***

In particular, on putting $r = 2,3\&4$ and noting that $\mu_1 = 0$, we get
$\mu_2' = \mu_2 + {\mu_1'}^2$
$\mu_3' = \mu_3 + 3\mu_2{\mu_1'}^2 + {\mu_1'}^3$
$\mu_4' = \mu_4 + 4\mu_3\mu_1' + 6\mu_2{\mu_1'}^2 + {\mu_1'}^4$

Moments about the Origin : The rth moment about the origin $\nu_r$is defined as
$\nu_r = \frac{1}{N}\sum_{i=1}^{n} f_i{x_i}^r, r = 0,1,2 \ldots\ldots$....where, $N = \sum_i f_i$
Putting$r = 0,1,2,..$

$$\nu_0 = 1, \nu_1 = \bar{x}, \nu_2 = \frac{1}{N}\sum_{i=1}^{n} f_i x_i^2$$

Relation Between Moments about Origin and about Mean:

$$\nu_r = \frac{1}{N}\sum_{i=1}^{n} f_i\, x_i^r ; r = 0,1,2 \dots\dots\dots$$

$$=\frac{1}{N}\sum_{i=1}^{n} f_i\,[(x_i - A)^r + \binom{r}{1}(x_i - A)^{r-1}.A + \cdots A^r]$$

If we take, $A = \bar{x}$ (for μ_r) then

$$\nu_r = \mu_r + \binom{r}{1}\mu_{r-1}\bar{x} + \binom{r}{2}\mu_{r-2}\bar{x}^2 + \cdots A^r \dots\dots\dots\dots\dots\dots(5.3.7)$$

$$\nu_1 = \bar{x}$$

$$\nu_2 = \mu_2 + \bar{x}^2$$

$$\nu_3 = \mu_3 + 3\mu_2\bar{x}^2 + \bar{x}^3$$

$$\nu_4 = \mu_4 + 4\mu_3\bar{x} + 6\mu_2\bar{x}^2 + \bar{x}^4.$$

Sheppard's Corrections for Moments: While computing moments for frequency distribution with class intervals, we take variable x as the mid-point of class-intervals which means that we have assumed the frequencies concentrated at the mid-points of class-intervals.

The above assumption is true when the distribution is symmetrical and the number of class intervals is not greater than $\frac{1}{20}^{th}$ of the range, otherwise the computation of moments with have certain error called grouping error. This error is corrected by the following formulae given by W.F.Sheppard. $\mu_2(\text{corrected}) = \mu_2 - \frac{h^2}{12}$. $\mu_4(\text{corrected}) = \mu_4 - \frac{1}{2}h^2\mu_2 + \frac{7}{240}h^4$

where h is the width of class-interval while μ_1 and μ_3 require no correction. These formulae are known as Sheppard's Corrections.

Karl Pearson's β and γ Coefficients: Karl Pearson defined the following four coefficients, based upon the first four moments about mean:

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3}, \gamma_1 = +\sqrt{\beta_1} \text{and} \beta_2 = \frac{\mu_4}{\mu_2^2}, \gamma_2 = \beta_2 - 3 \dots\dots\dots\dots\dots\dots(5.3.8)$$

The practical use of these coefficients is to measure the skewness and kurtosis of a frequency distribution. These coefficients are pure numbers independent of units of measurement.

- The mean, which indicates the central tendency of a distribution.
- The second moment is the variance, which indicates the width or deviation.

Symmetrical Distribution.

A distribution is said to be symmetrical when the distribution on either side of the mean is a mirror image of theother. In a symmetrical distribution, mean = median = mode.If a distribution is non-symmetrical, it is said to be skewed.

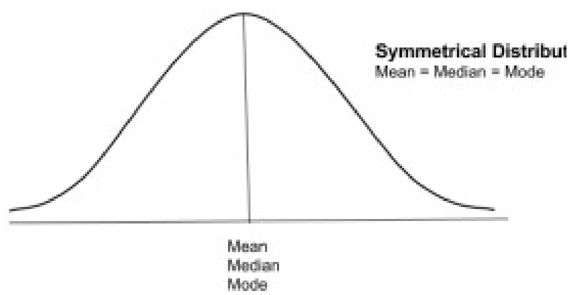


Fig 5.3.1

Skewness. Skewness is a measure of symmetry, or more precisely, the lack of symmetry in a frequency distribution.Skewness is positive if the longer tail of the distribution lies towards the right and negative if it lies towards the left.

Moment coefficient of skewness = $\frac{\mu_3}{\sqrt{\mu_2^3}}$

- If arithmetic mean < Mode (negative skewed).
- If arithmetic mean > Mode (positive skewed).
- If Sum of frequencies of values less than mode = sum of frequencies greater than mode it implies no skewness.

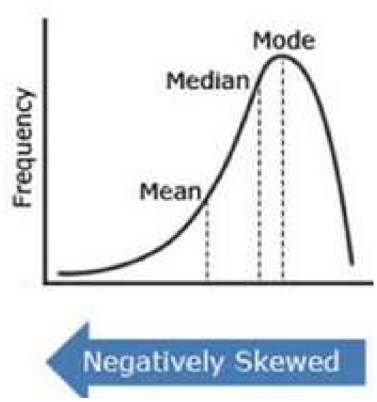


Fig.5.3.2

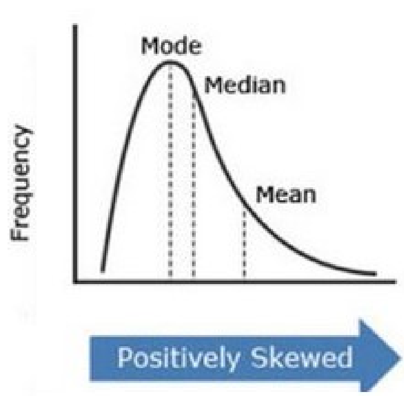


Fig.5.3.3.

Kurtosis. Kurtosis is a statistical measure used to describe a characteristic of a dataset. When normally distributed data is plotted on a graph, it generally takes the form of an upside down bell. This is called the bell curve. The plotted data that are furthest from the mean of the data usually form the tails on each side of the curve. Kurtosis indicates how much data resides in the tails. The relative flatness of the top is called kurtosis.

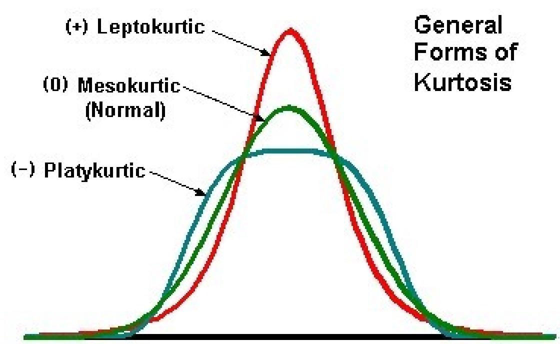


Fig.5.3.4.

Difference between skewness and kurtosis?

Skewness and kurtosis are both important measures of a distribution's shape. Skewness measures the asymmetry of a distribution. Kurtosis measures the heaviness of a distribution's tails relative to a normal distribution.

Measure of Kurtosis: The measure of kurtosis is denoted by β_2 and is defined as:

$$\beta_2 = \frac{\mu_4}{\mu_2^2}$$

- If $\beta_2 > 3$, the distribution is leptokurtic ($\gamma_2 > 0$).
- If $\beta_2 < 3$, the distribution is platykurtic ($\gamma_2 < 0$).
- If $\beta_2 = 3$, the distribution is mesokurtic ($\gamma_2 = 0$).

5.4.MATHEMATICAL EXPECTATION:-

The expected value of a discrete random variable is weighted average of all possible values of the random variable, where the weights are the probabilities associated with the corresponding values. The notation is $E(X)$and $E[X]$. Another popular notation is μ_X, whereas $< X >, < X >_{av}$, and are commonly used in physics and M(*X*) in Russian-language literature. The mathematical expression for computing the expected value of a discrete random variable X with probability mass function $(p.m.f.)f(x)$ is given below:

$E(X) = \sum_x xf(x)$(for discrete random variable) **……………*(5.4.1).***

The mathematical expression for computing the expected value of a continuous random variable X with probability density function $(p.d.f.)f(x)$ is, however, as follows:
$E(X) = \int_{-\infty}^{+\infty} xf(x)dx$,(for continuous random variable) **……*(5.4.2).***
provided the right, hand integral ***(5.4.1)*** and ***(5.4.2)*** is absolutely convergent,

i.e., provided $\int_{-\infty}^{+\infty}|xf(x)|dx = \int_{-\infty}^{+\infty}|x|f(x)dx < \infty \dots \dots (\mathbf{5.4.3})$
or $\sum_x|xf(x)| = \sum_x|x|\, f(x) < \infty \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots (\mathbf{5.4.4})$

Remark5.4.1: It should be clearly understood that although X has an expectation only if L.H.S. in ***(5.4.3)*** or ***(5.4.4)*** exists, i.e., converges to a finite limit, its value is given by (***5.4.3***) or ***(5.4.4).***
Remark 5.4.2:$E(X)$ exists if $E|X|$ exists.
Remark 5.4.3: Var$X$= $P(A)P(\bar{A})$.

Expected value of function of a random variable:

Consider a random variable with $p.d.f\ (p.m.f.)f(x)$ and distribution function $F(x)$. If $g(.)$ Is a function such that $g(X)$ is a random variable and $E[g(X)]$ exists (i.e., is defined), then
$E(X) = \int_{-\infty}^{+\infty} g(x)f(x)dx$,(for continuous random variable) ... $(\mathbf{5.4.5})$
$E(X) = \sum_x g(x)f(x)$(for discrete random variable) ... ... $(\mathbf{5.4.6})$

Remark 5.4.5:
μ_r'(about origin) = $E(X^r)$. μ_1'(about origin) = $E(X)$ and μ_2'(about origin) = $E(X^2)$.
Mean = $\bar{x}$ = μ_1'(about origin) = $E(X)$and $\mu_2 = \mu_2' - {\mu_1'}^2 = E(X^2) - \{E(X)\}^2$,
$\mu_2 = E[X - E(X)]^2 = \int_{-\infty}^{\infty}(x-\bar{x})^2 f(x)dx$**………………*(5.4.6a)***

Remark 5.4.6: $E(c) = c$ **..……………………………(5.4.7b)**

Properties of Expectation:

- If X and Y are random variables, then $E(X+Y) = E(X) + E(Y), \dots \dots \dots \dots \dots \dots .. (\mathbf{5.4.7})$
  Provided all the expectations exist.
- If X and Y are random variables, then $E(XY) = E(X).E(Y), \dots \dots \dots \dots \dots \dots .. (\mathbf{5.4.8})$
- If X is a random variable and $'a'$ is constant, then
  $(i) E[a\psi(X)] = aE[\psi(X)] \dots \dots \dots \dots \dots \dots \dots (\mathbf{5.4.9})$

$(i) E[\psi(X) + a] = E[\psi(X)] + a$ ... ... **(5.4.10)**

- If X is a random variable and a and b are constants, then $E[aX + b] = aE[X] + b$,
- If X is a random variable thus $E(X^2) \neq [E(X)]^2$ ... **(5.4.11)**
- $E[g(X)] = g\{E(X)\}$. ... ... **(5.4.12)**

- Let X_1, X_2, X_3 ... ... X_n be any n random variables and if a_1, a_2, a_3 ... ... a_n are n constants, then $E(\sum_{i=1}^{n} a_i X_i) = \sum_{i=1}^{n} a_i E(X_i)$ ... ... **(5.4.13)** provided all the expectations exist.
- If $X \geq 0$ then $E(X) \geq 0$ ....**(5.4.14)**
- If X and Y are two random variables such that $Y \leq X$, then $E(Y) \leq E(X)$, provided all the expectations exist**......*(5.4.15)***
- $|E(X)| \leq E|X|$, provided the expectations exist**......*(5.4.16)***
- If μ_r' exists, then μ_s' exists for all $1 \leq s \leq r$. Mathematically, if $E(X^r)$ exists, then $E(X^s)$ exist for all $1 \leq s \leq r, i.e., E(X^r) < \infty \Longrightarrow E(X^s) < \infty \forall 1 \leq s \leq r, i.e.,$**...(5.4.17)**
- If X and Y are two independent random variables, then $E[h(X).k(Y)] = E[h(X)]E[k(Y)]$**................(5.4.18).** Where $h(.)$ is a function of X alone and $k(.)$ is a function of Y alone, provided expectations on both side exist.

Cauchy-Schwartz Inequality . If X and Y are two random variables taking real values, then

$$[E(XY)]^2 \leq E(X^2).E(Y^2)] \text{..............................} \mathbf{(5.4.19)}.$$

Proof. Let us consider a real valued function of the real variable t, defined by

$$Z(t) = E(X + tY)^2$$

which is always non-negative,
since$(X + tY)^2 \geq 0$, for all real X , Y and t.
Thus

$$Z(t) = E(X + tY)^2 \geq 0 \forall t.$$

It implies

$$Z(t) = E[X^2 + 2tXY + t^2Y^2]$$

$$= E[X^2] + 2tE[XY] + t^2E[Y^2] \geq 0, \text{for all } t \text{......} \mathbf{(5.4.20)}.$$

Equation $(\mathbf{5.4.20})$ is a quadratic expression in $'t'$.

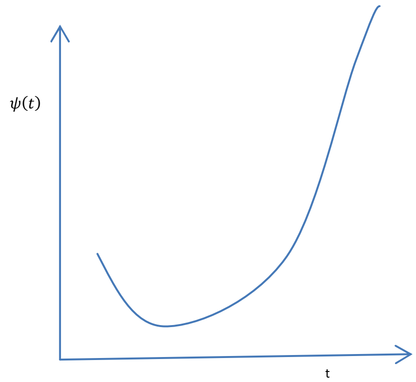


Fig 5.4.1

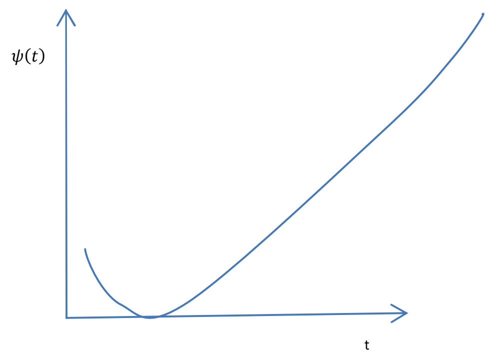


Fig 5.4.2

If quadratic expression $\psi(t) = At^2 + Bt + C \geq 0$ for all t, implies that the graph of the function $\psi(t)$ either touches the $t-$ axis at only one point or not at all, as exhibited in the above diagrams. This is equivalent to saying that the discriminant of the function $\psi(t)$, $viz., B^2 - 4AC \leq 0$, since the condition $B^2 - 4AC > 0$ implies that the function $\psi(t)$ has two distinct real roots which is contradiction to the fact that $\psi(t)$ meets the $t-$ axis either at only one point or not at all. Using the result, we get from equation (5.4.20),

$$4. E[XY\text{-}4. E[X^2]. E[Y^2] \leq 0 \Longrightarrow [E(XY)]^2 \leq E(X^2). E(Y^2)]$$

- Standard Deviation ≥ Mean Deviation**……………..(5.4.21).**

Continuous Convex Function. A continuous function $g(x)$ on the interval I is convex if for every $x_1$and x_2, $\frac{(x_1+x_2)}{2} \in I$.

$g\left(\frac{x_1+x_2}{2}\right) \leq \frac{1}{2}g(x_1) + \frac{1}{2}g(x_2)$…………………………………..**(5.4.22)**.

Jenson's Inequality. If g is continuous and convex function on the interval I, and X is a random variable whose values are in I with probability 1, then

$E[g(X)] \geq g[E(X)]$, provided the expectations exist**…….(5.4.22).**

Variance. The term variance refers to a statistical measurement of the spread between numbers in a data set. More specifically, variance measures how far each number in the set is from the mean (average), and thus from every other number in the set. If X is a random variable, then

Var$X$ = $E(X^2) - [E(X)]^2$.......................................***(5.4.19)***
Var$X$= Var($\boldsymbol{x}$) $= E(X-\mu)^2 = \sum px^2 - \mu^2$ ... ... $(\mathbf{5.4.20})$
Var$X$ = σ^2...............................***(5.4.21)***
Standard Deviation = $\sigma = \sqrt{\frac{35}{6}}$......................***(5.4.22)***

- If X is a random variable, then Var$(aX+b) = a^2V(X)$, where a and b are constants.
- If $b=0$, then Var$(aX) = a^2V(X)$, Variance is not independent of change of scale.
- If $a=1$, then Var$(X+b) = V(X)$.

5.5.MOMENT GENERATING FUNCTION:-

The moment generating function($m.g.f.$) of a, random variable X (about origin) having the probability function f(x) is given by:

$$M_X(t) = E(e^{tX})$$
$$= \begin{cases} \int e^{tX} f(x)dx, (for\ continuous\ probability\ distribution) \\ \sum_x e^{tX} f(x), (for\ discrete\ probability\ distribution) \end{cases} \quad \ldots (\mathbf{5.5.1})$$

the integration or summation being extended to the entire range of x, t being the real parameter and it is being assumed that the right-hand side of $(\mathbf{5.6.1})$ is absolutely convergent for some positive number h such that $-\mathrm{h} < t < h$. Thus

$$M_X(t) = E(e^{tX}) = E\left(1 + tX + \frac{t^2X^2}{2!} + \cdots \frac{t^rX^r}{r!} + \cdots\right)$$
$$= 1 + tE(X) + \frac{t^2}{2!}E(X^2) + \cdots \frac{t^r}{r!}E(X^r) + \cdots$$
$$= 1 + t\mu_1' + \frac{t^2}{2!}\mu_2' + \cdots \frac{t^r}{r!}\mu_r' + \cdots \quad (5.5.2)$$
$$= \sum_{r=0}^{\infty} \frac{t^r}{r!}\mu_r' \quad \ldots\ldots\ldots\ldots \mathbf{\mathit{(5.5.3)}}$$

$$\mu_r' = E(X^r) = \begin{cases} \int x^r f(x)dx, (for\ continuous\ probability\ distribution) \\ \sum_x x^r p(x), (for\ discrete\ probability\ distribution), \end{cases} \quad ..\ \mathbf{\mathit{(5.5.4)}}$$

is the rth moment of X about origin. Therefore the coefficient of $\frac{t^r}{r!}$ in $M_X(t)$ gives μ_r' (about origin). Since $M_X(t)$ generates moments, it is known as *moment generating function.* Differentiating ***(4.6.2)*** with respect to tr times and then putting $t = 0$, we get,

$$\mu_r' = \left|\frac{d^r}{dt^r}\{M_X(t)\}\right|_{t=0} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots..\mathbf{(5.5.5)}$$

$$M_X(t)(\text{about}X = a) = E\left(e^{t(X-a)}\right) = E\left(1 + t(X-a) + \frac{t^2(X-a)^2}{2!} + \cdots\frac{t^r(X-a)^r}{r!} + \cdots\right)$$

$$= 1 + t\mu_1' + \frac{t^2}{2!}\mu_2' + \cdots\frac{t^r}{r!}\mu_r' + \cdots\mathbf{(5.5.6)}.$$

Where $\mu_r' = E\{(X-a)^r\}$, is the rth moment about the point $X = a$.

Limitations of Moment Generating Function:

- A random variable X may have no moments although its $m.g.f.$ exists.
- A random variable X can have $m.g.f.$ and some (or all) moments, yet the
  $m.g.f.$ does not generate the moments.
- A random variable Xcan have all or some moments, but $m.g.f.$ does not exist except
  perhaps at one point.

Properties of Moment Generating Function:

- $M_{cX}(t) = M_X(ct), c$ being a constant**……………..**$\mathbf{(5.5.7)}$.
- The moment generating function of the sum of a number of independent random variables is equal to the product of their respective moment generating functions.
- **Uniqueness Theorem of Moment generating Function:**The moment generating function of a distribution, if it exists, uniquely determines the distribution. This implies that corresponding to a given probability distribution, there is only one $m.g.f.$ (provided it exists) and corresponding to a given $m.g.f.$ there is only one probability distribution. Hence $M_X(t) = M_Y(t) \Rightarrow X$ & Y are identically distributed.
- Effect of change of origin and scale on moment generating function. Let us transform X to the new variable U by changing both the origin and scale in X as follows:
  $U = \frac{X-a}{h}$,where$a$ and h are constants.
  The moment generating function of U(about origin) is given by:
  $$M_U(t) = E(e^{tU}) = E[\exp\{t(x - a/h)\} = E\left[e^{tX/}\ e^{-at/h}\right]$$
  $$= e^{-at/h}E\left(e^{tX/h}\right)$$
  $$= e^{-at/h}\ E\left(e^{Xt/h}\right) = e^{-at/h}M_X(t/h), \ldots\ldots\ldots\ldots\ldots\mathbf{(5.5.8)}.$$
  Where $M_X(t)$ is the moment generating function of X about origin. In particular, if we take
  $a = E(X) = E[\mathrm{X}] = \mu$ and $h = \sigma_X = \sigma$, then $U = \frac{X-E(X)}{\sigma_X} = \frac{X-\mu}{\sigma} = Z$, is known as a standard variate.
  Thus $m.g.f$ of a standard variate Z is given by

$$M_U(t) = e^{-\mu t/\sigma} M_X(t/\sigma)\ldots\ldots\ldots\ldots\ldots(\mathbf{5.5.9}).$$

Remark 4.6.1. $E(Z) = E\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma}E(X-\mu) = \frac{1}{\sigma}\{E(X)-\mu\} = \frac{1}{\sigma}\{E(X)-\mu\} = \frac{1}{\sigma}\{\mu-\mu\}$= 0 and$V(Z) = V\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma^2}V(X-\mu) = \frac{1}{\sigma^2}V(X) = 1$. Therefore $E(Z) = 0$, $V(Z) = 1$, i.e mean and variance of a standard variate are 0 and 1 respectively.

Cumulants: Cumulants generating function$K_X(t) = log_e M_x(t)$, provided the right-hand side can be expanded as a convergent series in powers of t.

Mean=k_1
$\mu_2 = k_2$ =Variance
$\mu_3 = k_3$
$\mu_4 = k_4 + 3k_2^2$

5.6. CHARACTERISTIC FUNCTION:-

In probability theory and statistics, the **characteristic function** of any real-valued random variable completely defines its probability distribution. If a random variable admits a probability density function, then the characteristic function is the Fourier transform of the probability density function. Thus it provides an alternative route to analytical results compared with working directly with probability density functions or cumulative distribution functions. The characteristic function is a way for describing a random variable.
The characteristic function

$$\emptyset_X(t) = E\left(e^{itX}\right) = \begin{cases} \int e^{itx} f(x)dx \ (for\ continuous\ probability\ distribution) \\ \sum_x e^{itx}\, p(x)(for\ discrete\ probability\ distribution) \end{cases} \quad \ldots\textit{(5.6.1)}$$

If $F_X(x)$ is the distribution function of a continuous random variable X, then

$$\emptyset_X(t) = \int_{-\infty}^{\infty} e^{itx} dF\,(x) \ldots\ldots\ldots\ldots\ldots(\mathbf{5.6.2})$$

$$|\emptyset(t)| \le 1 \ldots\ldots\ldots\ldots\ldots(\mathbf{5.6.3}).$$

Characteristic function $\emptyset_X(t)$ always exists.

Remark 5.6.1.

$$\emptyset_X(t) = E\left(e^{itX}\right) = E\left[1 + itX + \frac{(it)^2X^2}{2!} + \cdots \frac{(it)^r X^r}{r!} + \cdots\right]$$

$$= 1 + tE(X) + \frac{(it)^2}{2!}E(X^2) + \cdots \frac{(it)^r}{r!}E(X^r) + \cdots$$

$= 1 + it\mu_1' + \frac{(it)^2}{2!}\mu_2' + \cdots \frac{(it)^r}{r!}\mu_r' + \cdots$ ***(5.6.4)***

where $\mu_r' = E(X^r)$, is the rth moment of X about origin.

Therefore $\mu_r' =$ Coefficient of $\frac{(it)^r}{r!}$ in $\emptyset_X(t)$ … … … … … . . $(\mathbf{5.6.5})$

- $m.g.f.$ and characteristic function $\emptyset(t)$ both generate moments.
- Cumulant generating function $K_X(t)$ in terms of $\emptyset_X(t)$ is given by:
  $$K(t) = \sum_{r=1}^{\infty} \frac{(it)^r}{r!} k_r,$$
  wherek_r =Coefficient of $\frac{(it)^r}{r!}$ in $K_X(t)$, is the rth cumulant.

Properties of Moment Generating Function:

- For all real t,
  (i) $\emptyset(0) = 1$ … … … … … … … . $(\mathbf{5.6.6})$
  (ii) $|\emptyset(t)| \leq 1$ … … … … … … … . $(\mathbf{5.6.7})$
- $\emptyset(t)$is continuous and uniformly continuous for all real t.
- $\emptyset_X(-t)$and$\emptyset_X(t)$ are conjugate functions, i.e., $\emptyset_X(-t) = \overline{\emptyset_X(t)}$, where $\bar{a}$ is the complex conjugate of a….. $(\mathbf{5.6.8})$.
- If the distribution function of a random variable X is symmetrical about zero, i.e., if
  $$1 - F(x) = F(-x) \Rightarrow f(-x) = f(x),$$
  then$\emptyset_X(t)$ is real valued and even function of t……. $(\mathbf{5.6.9})$.
- IfX in some random variable with characteristic function $\emptyset_X(t)$, and if $\mu_r' = E(X^r)$ exists, then$\mu_r' = (-1)^r \left|\frac{\partial}{\partial t^r}\emptyset(t)\right|_{t=0}$……………………………… $(\mathbf{5.6.10})$.
- $\emptyset_{cX}(t) = \emptyset_X(ct)$, $c$being a constant……………$(\mathbf{5.6.11})$.
- The characteristic function of the sum of a number of independent random variables is equal to the product of their respective characteristic functions…………$(\mathbf{5.6.12})$.
- **Uniqueness Theorem of characteristic function:**Characteristic function uniquely determined thedistribution, i.e., a necessary and sufficient condition for two distributions with$p.d.f.'s$ (provided it exists) $f_1(.)$ and $f_2(.)$ to be identical is that their characteristic functions $\emptyset_1(t)$ and $\emptyset_2(t)$ are identical…………$(\mathbf{5.6.13})$.
- Effect of change of origin and scale on Characteristic Function. Let us transform X to the new variable U by changing both the origin and scale in X as follows:
  $U = \frac{X-a}{h}$,where$a$ and $h$being constants, then $\emptyset_U(t) = e^{-iat/h}\emptyset_X(t/h)$

In particular, if we take
$a = E(X) = E[X] = \mu$ and $h = \sigma_X = \sigma$, then $Z = \frac{X-E(X)}{\sigma_X} = \frac{X-\mu}{\sigma}$, is given by

$M_U(t) = e^{-i\mu t/\sigma}\phi_X(t/\sigma)$ …. **(5.6.14).**

- The characteristic function of a real-valued random variable always exists, since it is an integral of a bounded continuous function over a space whose measure is finite. A characteristic function is uniformly continuous on the entire space. characteristic function is infinitely differentiable

CHECK YOUR PROGRESS

Q1. Which formula represents the 2^{nd} moment around the mean?
(i)$\sum(x_i - \mu_x)^2$ (ii)$\sum x_i$ (iii) $\sum(x_i - \mu_x)^3$ (iv)$\sum(x_i - \mu_x)^4$
Q2. The r^{th} moment of a random variable X is given by:
(i) $E(X^r)$ (ii) $E(X^2)$ (iii)$E(X^{r-1})$ (iv)$None\ of\ these$
Q3. Let$X$ is a real valued random variable with $E[X]$ and $E[X^2]$ denoting the mean values of X and X^2 respectively. The relation which always holds:
(i) $E(X^2) \geq (E[X])^2$(ii) $E(X^2) \leq (E[X])^2)$ (iii) $E(X^2) \leq (E[X])^2$(iv)None of these
Q4. Which inequality is correct?
(i)Var$X$ = $E(X^2) - [E(X)]^2$ (ii))Var$X$>$E(X^2) - [E(X)]^2$
(iii) **)** Var$X$<$E(X^2) - [E(X)]^2$ (iv) None of these
Q5. Mean of a constant $'a'$ is……………………..
Q6. Variance of a constant $'a'$ … ...…………………..
Q7. If X is a random variable such that $P(a \leq X \leq b) = 1$, then $E(X)$ and Var (X) exist. True\False
Q8.The moment-generating function of a real-valued distribution does not always exist. True\False
Q9.Characteristic function does not exists for every random variable
True\False
Q10.$M_{7X}(t) \neq M_X(7t)$.True\False

5.7.SOLVED EXAMPLES:-

Example 5.7.1. Let X be a discrete random variable having probability mass function

$$p_X(x) = \begin{cases} 1/2\ if\ x = 1 \\ 1/3\ if\ x = 2 \\ 1/6\ if\ x = 3 \\ 0\ otherwiise \end{cases}$$

Find the third moment of X.

Solution. Third moment = $E(X^3) = \sum(x^3) = \sum x^3\, p_X(x)$

$$= \frac{1}{2}.1^3 + \frac{1}{3}.2^3 + \frac{1}{6}.3^3$$

$$= \frac{1}{2} + \frac{8}{3} + \frac{27}{6} = \frac{23}{3}$$

Example 5.7.2. Let X be a discrete random variable having probability mass function

$$p_X(x) = \begin{cases} 3/4 \; if \; x = 1 \\ 1/4 \;\; if \; x = 2 \\ 0 \; otherwise \end{cases}$$

Find the third central moment of X.

Solution. Third moment = $E(X) = \mu_x = \sum x\, p_X(x)$

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

The third central moment of X can be computed as follows:

$E(X - \mu_x)^3 = \sum \left(x - \frac{5}{4}\right)^3 p_X(x) = \left(1 - \frac{5}{4}\right)^3 \times \frac{3}{4} + \left(2 - \frac{5}{4}\right)^3 \times \frac{1}{4} = \left(-\frac{1}{4}\right)^3 . \frac{3}{4} + \left(\frac{3}{4}\right)^3 . \frac{1}{4} = \frac{3}{32}.$

Example 5.7.3. Find the first four moments for the following individual series:

x	3	6	8	10	18

Solution.

Calculation of Moments

S.No.	$\boldsymbol{x}$	$\boldsymbol{x - \bar{x}}$	$\boldsymbol{(x - \bar{x})^2}$	$\boldsymbol{(x - \bar{x})^3}$	$\boldsymbol{(x - \bar{x})^4}$
1.	3	-6	36	-216	1296
2.	6	-3	9	-27	81
3.	8	-1	1	-1	1
4.	10	1	1	1	1
5.	18	9	81	729	6561
$n = 5$	$\sum x = 45$	$\sum(x - \bar{x}) = 0$	$\sum(x - \bar{x})^2 = 128$	$\sum(x - \bar{x})^3 = 486$	$\sum(x - \bar{x})^4 = 7940$

Now, $\bar{x} = \frac{\sum x}{n} = \frac{45}{5} = 9.$

Therefore $\mu_1 = \frac{\sum(x - \bar{x})}{n} = \frac{0}{5} = 0,$

$$\mu_2 = \frac{\sum(x - \bar{x})^2}{n} = \frac{128}{5} = 25.6,$$

$$\mu_3 = \frac{\sum(x - \bar{x})^3}{n} = \frac{486}{5} = 97.2,$$

$$\mu_4 = \frac{\sum(x - \bar{x})^4}{n} = \frac{7940}{5} = 1588.$$

Example 5.7.4. Calculate $\mu_1, \mu_2, \ \mu_3, \mu_4$ for the following frequency distribution:

Marks	5-15	15-25	25-35	35-45	45-55	55-65
No. of students	10	20	25	20	15	10

Solution.

$Marks$	$No. of\ studen\ (f)$	$Mid-point\ (x)$	fx	$x-\bar{x}$	$f(x-\bar{x})$	$f(x-\bar{x})^2$	$f(x-\bar{x})^3$	$f(x-\bar{x})^4$
5-15	10	10	100	-24	-240	5760	-138240	3317760
15-25	20	20	400	-14	-280	3920	-54880	768320
25-35	25	30	750	-4	-100	400	-1600	6400
35-45	20	40	800	6	120	720	4320	25920
45-55	15	50	750	16	240	3840	61440	983040
55-65	10	60	600	26	260	6760	175760	4569760
	$N = 100$		$\sum fx = 340($		$\sum f(x-\bar{x})=0$	$\sum f(x-\bar{x})^2 = 21400$	$\sum f(x-\bar{x})^3 = 46800$	$\sum f(x-\bar{x})^4 = 9671200$

Now, $\bar{x} = \frac{\sum fx}{N} = \frac{3400}{100} = 34.$

Therefore $\mu_1 = \frac{\sum f(x-\bar{x})}{N} = \frac{0}{100} = 0,$

$$\mu_2 = \frac{\sum f(x-\bar{x})^2}{N} = \frac{21400}{100} = 214,$$

$$\mu_3 = \frac{\sum f(x-\bar{x})^3}{N} = \frac{46800}{100} = 468,$$

$$\mu_4 = \frac{\sum f(x-\bar{x})^4}{N} = \frac{9671200}{100} = 96712.$$

Example 5.7.5. The first three moments of a distribution about the value $'2'$ of the variable are $1, 16$ and -40. Show that the mean is 3, variance is 15 and $\mu_3 = -86$.

Solution. We have $A = 2, \mu_1' = 1, \mu_2' = 16$ and $\mu_3' = -40$.

$\mu_1' = \bar{x} - A \Rightarrow \bar{x} = \mu_1' + A = 1 + 2 = 3.$

Variance $= \mu_2 = \mu_2' - {\mu_1'}^2 = 16 - (1)^2 = 15.$

$$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2{\mu_1'}^3 = -40 - 3(16)(1) + 2(1)^3$$
$$= -40 - 48 + 2 = -86$$

Example 5.7.6. Calculate the variance and third central moment from the following data:

x_i	0	1	2	3	4	5	6	7	8
f_i	1	9	26	59	72	52	29	7	1

Solution. **Calculation of Moments**

x	f	$u = \frac{x-a}{h}$ $A = 4, h = 1$	fu	fu^2	fu^3
0	1	-4	-4	16	-64
1	9	-3	-27	81	-243
2	26	-2	-52	104	-208
3	59	-1	-59	59	-59
4	72	0	0	0	0
5	52	1	52	52	52
6	29	2	58	116	232
7	7	3	21	63	189
8	1	4	4	16	64
	$N = \sum f = 256$		$\sum fu = -7$	$\sum fu^2 = 507$	$\sum fu^3 = -37$

Now, moments about the point $x = A = 4$ are

$$\mu_1' = \left(\frac{\sum fu}{N}\right) h = \frac{-7}{256} = -0.02734$$

$$\mu_2' = \left(\frac{\sum fu^2}{N}\right) h^2 = \frac{507}{256} = 1.9805$$

$$\mu_3' = \left(\frac{\sum fu^3}{N}\right) h^3 = \frac{-37}{256} = -0.1445$$

moments about mean

$$\mu_1 = 0, \mu_2 = \mu_2' - {\mu_1'}^2 = 1.9805 - (-.02734)^2 = 1.97975$$

Variance = 1.97975.

Also,

$$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2{\mu_1'}^3 = (-0.1445) - 3(1.9805)(-.02734) + 2(-.02734)^3 = 0.0178997.$$

Third central moment = 0.0178997.

Example 5.7.7. The first four moments of a distribution about the mean are $0, 100, -7$ and 35000. Discuss the kurtosis of the distribution.

Solution. $\mu_1 = 0, \quad \mu_2 = 100, \quad \mu_3 = -7, \mu_4 = 35000.$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{35000}{(100)^2} = 3.5 > 3.$$

Therefore the distribution is leptokurtic.

Example 5.7.8. The first four moments of a distribution about the value $'5'$ of the variable are $2, 20, 40$ and 50. Calculate the moment coefficient of skewness.

Solution. We have $A = 5$, $\mu_1' = 2, \mu_2' = 20, \mu_3' = 40, \mu_4' = 50$

$$\mu_2 = \mu_2' - {\mu_1'}^2 = 20 - (2)^2 = 16$$

Also $\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2{\mu_1'}^3 = 40 - 3(2)(20) + 2(2)^3 = 40 - 120 + 16 = -64.$

Moment coefficient of skewness $= \frac{\mu_3}{\sqrt{\mu_2^3}} = \frac{-64}{\sqrt{(16)^3}} = \frac{-6}{64} = -1$

Example.5.7.9.

i A pair of two coins is tossed, what is the expected value?

ii A pair of dice is thrown together, find the expected value.

Solution.

i Expected value or mean value = $E(X) = \mu = \sum_{i=1}^{n} p_i\, x_i$ (Here X is a discrete random variable. In tossing of two coins, probability distribution is represented in tabular as follows:

	0	1	2
$P(x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

Therefore, $E(X) = \frac{1}{4} \times 0 + \frac{1}{2} \times 1 + \frac{1}{4} \times 2 = 1.$

As the probability of getting no head, one head and two heads is respectively $\frac{1}{4}, \frac{1}{2}, \frac{1}{4}$.

(ii) In a throw of pair of dice the sum (X) is a discrete random variable which is an integer between 2 and 12 with the probabilities as given below.

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

Therefore Expected value =

$$E(X) = \mu$$
$$= \frac{1}{36}.2 + \frac{2}{36}.3 + \frac{3}{36}.4 + \cdots \ldots \ldots + \frac{2}{36}.11 + \frac{1}{36}.12$$
$$= \frac{252}{36} = 7$$

- The variance in each of the above cases is given by: Var$X$= $\sum px^2 - \mu^2$

In the tossing of two coins, we have

$$\sum p.x^2 = \frac{1}{4}.(0)^2 + \frac{1}{2}.(1)^2 + \frac{1}{4}.(2)^2 = \frac{3}{2}$$

Therefore Variance= $\sum px^2 = \frac{1}{36}.4 + \frac{2}{26}.9 + \frac{3}{36}.16 + \frac{4}{36}.25 + \frac{5}{36}.36 + \frac{6}{36}.49 + \frac{5}{36}.64 + \frac{4}{36}.81 + \frac{3}{36}.100 + \frac{2}{36}.121 + \frac{1}{36}.144$

$$= \frac{1}{36}[4 + 9 + 16 + 25 + 36 + 49 + 64 + 81 + 100 + 121 + 144]$$

$= \frac{1}{36}[1974] = \frac{329}{6}$

Therefore Variance = $\sigma^2 = \frac{329}{6} - (7)^2 = \frac{35}{6}$.

Standard deviation = $\sigma = \sqrt{\frac{35}{6}}$.

Example.5.7.10. If X is a continuous random variable and K is a constant then prove that :

$$(i) V(x+k) = V(X), \qquad (ii) V(Xk) = k^2 V(X).$$

Solution. Var$X$ = $E(X^2) - [E(X)]^2$

Var$[X+k]$ = $E[(X+k)^2] - [E(X+k)]^2$.

$= E[X^2 + 2kX + k^2] - \{[E(X)]^2 + 2kE(X) + k^2\}$.

$= E[X^2] + 2kE(X) + k^2 - [E(X)]^2 - 2kE(X) - k^2$.

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

Therefore, $V(x+k) = V(X)$. Now, Var$[Xk]$= $E[(Xk)^2] - [E(Xk)]^2$.

$= E[k^2X^2] - [kE(X)]^2$

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

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

$= k^2 Var[X]$

Example.5.7.11. Let the random variable X have the distribution:

$$P(X=0) = P(X=2) = p; P(X=1) = 1-2p, for\ 0 \le p \le \frac{1}{2}.$$

For what p is that Var(X) a maximum?

Solution: In this example the random variable X takes the values 0, 1 and 2 with respective probabilities $p, 1-2p$ and $p, 0 \le p \le \frac{1}{2}$. Thus

$$E(X) = 0 \times p + 1 \times (1-2p) + 2 \times p = 1, E(X^2)$$
$$= 0 \times p + 1^2 \times (1-2p) + 2^2 \times p = 1 + 2p$$

Therefore, Var$X$ = $E(X^2) - [E(X)]^2 = 2p; 0 \le p \le \frac{1}{2}$.

Obviously, for $0 \le p \le \frac{1}{2}$, Var(X) is maximum when $p = \frac{1}{2}$, and $[Var(X)]_{max} = 2 \times \frac{1}{2} = 1$.

Example.5.7.12. Find the moment generating function of the exponential distribution.

$$f(x) = \frac{1}{c} e^{-\frac{x}{c}}; 0 \le x \le \infty, c > 0.$$

Hence find its mean and standard deviation.

Solution: Moment generating function about the origin is given by

$$M_x(t) = \int_0^\infty e^{tx} . \frac{1}{c} e^{-\frac{x}{c}} dx$$

$$= \frac{1}{c} \int_0^\infty e^{\left(t-\frac{1}{c}\right)x} dx$$

$$= \frac{1}{c}\left[\frac{e^{\left(t-\frac{1}{c}\right)x}}{\left(t-\frac{1}{c}\right)}\right]_0^\infty = (1-ct)^{-1} = 1 + ct + c^2t^2 + c^3t^3 + \cdots$$

Therefore $\nu_1 = \left[\frac{d}{dt} M_x(t)\right]_{t=0} = (c + 2c^2t + 3c^3t^2 + \cdots)_{t=0} = c$

$$\nu_2 = \left[\frac{d^2}{dt^2} M_x(t)\right]_{t=0} = 2c^2$$

Now, mean $\bar{x} = \nu_1 = c$
Variance $\mu_2 = \nu_2 - \bar{x}^2 = \nu_2 - {\nu_1}^2 = 2c^2 - c^2 = c^2$.
Therefore standard deviation $=\sqrt{\mu_2} = c$.

Example.5.7.13. Let $f_X(x) = \mu e^{-\mu x}$ where X be an exponential random variable with parameter μ. Find its characteristic function.

Solution: The characteristic function, $\emptyset_X(t) = \int_{x=0}^\infty \mu e^{-\mu x} e^{itx} dx$

$$= \frac{\mu}{\mu - it}.$$

We have evaluate the above integral essentially by pretending that $\mu - it$ is a real number.

5.8. SUMMARY:-

This unit the idea about "Expectation and moments". Here, we will introduce and discuss moment generating functions (MGFs). There are random variables for which the moment generating function does not exist on any real interval with positive length. If a random variable does not have a well-defined MGF, we can use the characteristic function defined.

5.9.GLOSSARY:-

i random variable
ii probability density function (p.d.f)
iii Mean
iv Variance
v Moments
vi Mathematical expectation
vii Moment generating function
viii Characteristic function

5.10.REFERENCE BOOKS:-

1. S. C. Gupta and V. K. Kapoor, (2020), *Fundamentals of mathematical statistics*, Sultan Chand & Sons.
2. Seymour Lipschutz and John J. Schiller, (2017), *Schaum's Outline: Introduction to Probability and Statistics*, McGraw Hill Professional.
3. J. S. Milton and J. C. Arnold , (2003*), Introduction to Probability and Statistics (4th Edition),* Tata McGraw-Hill.
4. https://www.wikipedia.org.

5.11. SUGGESTED READINGS:-

1. A.M. Goon,(1998), *Fundamental of Statistics 7th Edition),* 1998.
2. R.V. Hogg and A.T. Craig, (2002), *Introduction to Mathematical Statistics,* MacMacMillan, 2002.
3. Jim Pitman, (1993), *Probability,* Springer-Verlag.
4. https://archive.nptel.ac.in/courses/111/105/111105090.

5.12.TERMINAL QUESTIONS:-

TQ1. The first four moments of a distribution, about the value $'35'$ are $-1.8, 240, -1020$ and 144000. Find the values of $\mu_1, \quad \mu_2, \ \mu_3, \mu_4$.

TQ2. The following table represents the height of a batch of 100 learners. Calculate Kurtosis?

Height (cm)	59	61	63	65	67	69	71	73	75
No. of learners	0	2	6	20	40	20	8	2	2

TQ3. An urn contains 7white and 3 red balls. Two balls are drawn together, at random from this urn. Compute the probability that neither of them is white. Find also the probability of getting one white and one red ball. Hence compute the expected number of white balls drawn.

TQ4. Obtain the moment generating function at the random variable x having probability distribution

$$f(x) = \begin{cases} x \ for \ 0 < x < 1 \\ 2 - x \ for \ 1 \le x < 2 \\ 0 \ otherwise \end{cases} .$$ Also determine ν_1, ν_2 and μ_2.

TQ5. Let $f_X(x) = \frac{1}{\pi(1+x^2)}$ where X be anCauchy random variable.Find its characteristic function.

5.13. ANSWERS:-

Answer of Check your progress Questions:-

1) $\sum(x_i - \mu_x)^2$
2) $E(X^r)$
3) $E(X^2) \geq (E[X])^2$
4) Var$X = E(X^2) - [E(X)]^2$
5) a.
6) 0.
7) True
8) True
9) False
10) False

Answer of Terminal Questions:-

1) $\mu_1 = 0,\ \ \mu_2 = 236.76,\ \mu_3 = 264.36, \mu_4 = 141290.11.$
2) leptokurtic.
3) $\frac{21}{15}$.
4) $M_x(t) = 1 + t + t^2 + \cdots,\ \nu_1 = 1, \nu_2 = 2$ and $\mu_2 = 1$.
5) The characteristic function, $\emptyset_X(t) = e^{-|t|}$.

UNIT 6:- LAW OF LARGE NUMBERS

CONTENTS:

6.1.INTRODUCTION:-

In previous unit we have discussed about Mathematical expectation, Moment generating function and Characteristic function. In this unit first we explained about Chebychev's Inequality. This inequality is named after Russian mathematician Pafnuty Chebyshev, although it was first formulated by his friend and colleague Irénée-Jules Bienaymé. The theorem was first stated without proof by Bienaymé in ***1853*** and later proved by Chebyshev in ***1867***. The convergence in Probability is also explained here. The basic idea behind this type of convergence is that the probability of an "unusual" outcome becomes smaller and smaller as the sequence progresses. The idea of Weak Law of Large Numbers (W.L.L.N) is also discussed here. Some other concept of Large number is also explained.

6.2.OBJECTIVES:-

After studying this unit learner will be able to:

1. Describe the notion of Chebychev's Inequality.
2. Explain the concept of Convergence in Probability.
3. Defined the Weak Law of Large Numbers (W.L.L.N).
4. Apply the concept of Bernouli's Law of Large Numbers.
5. Summarize the Khinchin's Theorem,Borel-Cantlli Lemma (Zero-One Law) and Probability Generating Function (p.g.f).

6.3.CHEBYCHEV'S INEQUALITY:-

Chebyshev's inequality is a theory describing the maximum number of extreme values in a probability distribution.

Conditions for Chebyshev's inequality:

It requires only two minimal conditions:

i. That the underlying distribution have a mean.
ii. That the average size of the deviations away from this mean (as gauged by the standard deviation) not be infinite.

Definition. If X is a random variable with mean μ and variance σ^2, then for any positive number k, we have

$$P\{|X-\mu| \geq k\sigma\} \leq \frac{1}{k^2}$$

or $\{|X-\mu| < k\sigma\} \geq 1-\frac{1}{k^2}$ ***..............(6.3.1)***

Proof. Case (i). X is a continuous random variable. By definition,

$$\sigma^2 = \sigma_{X^2} - E\{X-E(X)\}^2 = E(X-\mu)^2$$

$= \int_{-\infty}^{\infty}(X-\mu)^2 f(x)dx$,where$f(x)$ is $p.d.f$ of$X$.

$$= \int_{-\infty}^{\mu-k\sigma}(X-\mu)^2 f(x)dx + \int_{\mu-k\sigma}^{\mu+k\sigma}(X-\mu)^2 f(x)dx + \int_{\mu+k\sigma}^{\infty}(X-\mu)^2 f(x)dx$$

$\geq \int_{-\infty}^{\mu+k\sigma}(X-\mu)^2 f(x)dx + \int_{\mu+k\sigma}^{\infty}(X-\mu)^2 f(x)dx$***...(6.3.2)***

Since, we know that
$x \leq \mu-k\sigma$and$x \geq \mu+k\sigma \Leftrightarrow |x-\mu| \geq k\sigma$ ... ... $(\mathbf{6.3.3})$
Therefore by using ***(6.3.2)*** and ***(6.3.3)*** we get,

$$\sigma^2 \geq k^2\sigma^2\left[\int_{-\infty}^{\mu+k} f(x)\,dx + \int_{\mu+k}^{\infty} f(x)\,dx\right]$$

$$\sigma^2 = k^2\sigma^2[P(X \leq \mu - k\sigma) + P(X \geq \mu + k\sigma)]$$

$$= k^2\sigma^2 . P\{|X - \mu| \geq k\sigma\}$$

(From **(5.3.3))**

This implies that $\{|X - \mu| \geq k\sigma\} \leq \mathbf{1/k^2}$ ... $(\mathbf{6.3.4})$

Also since $P\{|X - \mu| \geq k\sigma\} + P\{|X - \mu| < k\sigma\} = 1$, we get

$P\{|X - \mu| < k\sigma\} = 1 - P\{|X - \mu|\} \geq 1 - \{1/k^2\}$……***(6.3.5)***

Case (ii). IfX is a discrete random variable, the proof follows exactly similarly on replacing integration by summation.

Remark: In particular, if we take$k\sigma = c > 0$,then ***(6.3.2)* and *(6.3.3)*** give respectively

$P\{|X - \mu| \geq c\} \leq \frac{\sigma^2}{c^2}$and $P\{|X - \mu| < c\} \leq 1 - \frac{\sigma^2}{c^2}$

$P\{|X - \mu| \geq c\} \leq \frac{Var\,(X)}{c^2}$ and $P\{|X - \mu| < c\} \leq 1 - \frac{Var\,(X)}{c^2}$………***(6.3.6)***

Practical Example:

- Assume that an asset is picked from a population of assets at random. The average return of the population of assets is 12%, and the standard deviation of the population of assets is 5%. To calculate the probability that an asset picked at random from this population, which has a return less than 4% or greater than 20%, Chebyshev's inequality can be applied.

- Suppose we have sampled the weights of dogs in the local animal shelter and found that our sample has a mean of 20 pounds with a standard deviation of 3 pounds. With the use of Chebyshev's inequality, we know that at least 75% of the dogs that we sampled have weights that are two standard deviations from the mean. Two times the standard deviation gives us 2 x 3 = 6. Subtract and add this from the mean of 20. This tells us that 75% of the dogs have weight from 14 pounds to 26 pounds.

Use of the Inequality: If we know more about the distribution that we're working with, then we can usually guarantee that more data is a certain number of standard deviations away from the mean. For example, if we know that we have a normal distribution, then 95% of the data is two standard deviations from the mean. Chebyshev's inequality says that in this situation we know that *at least* 75% of the data is two standard deviations from the mean. As we can see in this case, it could be much more than this 75%.

The value of the inequality is that it gives us a "worse case" scenario in which the only things we know about our sample data (or probability distribution) is the mean and standard deviation. When we know nothing else about our data, Chebyshev's inequality provides some additional insight into how spread out the data set is.

6.3.1. GENERALISED FORM OF BIENAYME - CHEBYCHEV'S INEQUALITY:-

Let $g(X)$be a non-negative function of a random variable X. Then for every $k > 0$, we have, $P\{g(X) \geq k\} \leq \frac{E\{g(X)\}}{k}$**……….......(6.3.7)**

Proof. Here we shall prove the theorem for continuous random variable. The proof can be adapted to the case of discrete random variable on replacing integration by summation over the given range of the variable.

Let S be the set of all X, where $g(X) \geq k,\ i.e.,\quad S = \{x: g(x) \geq k\}$ then

$\int_S dF(x) \quad = \quad P\{g(X) \geq k\}$,………*(6.3.8)*

where$F(x)$ is the distribution function of X.

$$E\{g(X)\} = \int_{-\infty}^{\infty} g(x)dF(x) \geq \int_S g(x)dF(x) \geq kP\{g(x) \geq k\}$$

[Since on $S, g(x) \geq k$ and using *(6.3.8)]*

$P\{g(X)\} \geq k \leq \frac{E\{g(X)\}}{k}$………*(6.3.9)*

Remark 6.3.1. If we take$g(X) = \{X - E(X)\}^2 = \{X - \mu\}^2$and replace by $k^2\sigma^2$ in *(6.3.7),* we get

$$P\{(X-\mu)^2\} \geq \frac{E(X-\mu)^2}{k^2\sigma^2} = \frac{\sigma^2}{k^2\sigma^2} = \frac{1}{k^2}$$

$$\Rightarrow\ P\left\{|X-\mu| \geq k\sigma \leq \frac{1}{k^2}\right\} \text{… … …}\textit{(6.3.10).}$$

Which is Chebychev's inequality.

Remark 6.3.2.Markov's Inequality. Taking $g(X) = |X|$ in *(6.3.7)* we get, for any $k > 0$, $P\left\{|X| \geq k \leq \frac{E|X|}{k}\right\}$,which is Markov's inequality……………….*(6.3.11)*

The generalized form of Markov's inequality,

$$P(|X|^r \geq k^r) \leq \frac{E|X|^r}{k^r} \text{… … … … … … … …} (\mathbf{6.3.12})$$

Remark 6.3.3. If we assume the existence of only second – order moments of X, then we cannot do better than Chebychev's inequality. However, we can sometimes improve upon the results of Chebyshev's inequality if we assume the existence of higher order moments.

6.4.CONVERGENCE IN PROBABILITY:-

We shall now defined the concept of convergence, viz., convergence in probability or stochastic convergence. The idea that a sequence of essentially random or unpredictable events can sometimes be expected to settle down into a behaviour that is essentially unchanging when items far enough into the sequence are studied.

The different possible notions of convergence relate to how such behaviour can be characterized: two readily understood behaviours are that the sequence eventually takes a constant value, and that values in the sequence continue to change but can be described by an unchanging probability distribution.

A sequence of random variables $X_1, X_2, X_3, \ldots\ldots X_n \ldots\ldots\ldots$ is said to converge in probability to a constant a, if for any $\in > 0$,

$$\lim_{n\to\infty} P(|X_n - a| < \in) = 1 \ldots\ldots\ldots\ldots\ldots\ldots (\mathbf{6.3.13})$$

or its equivalent

$$\lim_{n\to\infty} P(|X_n - a| \geq \in) = 0 \ldots\ldots\ldots\ldots\ldots\ldots (\mathbf{6.3.14})$$

and we can written $X_n \xrightarrow{p} a, as\ n \to \infty$ ……***(6.3.15)***

If there exists a random variable X such that $X_n - a \xrightarrow{p} a$ as $n \to \infty$, then we say that the given sequence $\{X_n\}$ of random variables converges in probability to the random variable X.

Remark 6.4.1: If a sequence of constants $a_n - a$ as $n \to \infty$ then regarding the constant random variable having one-point distribution at that point, simply we write $a_n \xrightarrow{p} a$ as $n \to \infty$.

Remark 6.4.2: Although the concept of convergence in probability is basically different from that of ordinary convergence of sequence of numbers, it can easily verified that the following simple rules hold for convergence in probability as well.

If that $X_n \xrightarrow{p} \alpha\ Y_n \xrightarrow{p} \beta\ as\ n \to \infty$, then

i $X_n + Y_n \xrightarrow{p} \alpha \pm \beta\ as\ n \to \infty$,
ii $X_n Y_n \xrightarrow{p} \alpha\ \beta\ as\ n \to \infty$,
iii $\frac{X_n}{Y_n} \xrightarrow{p} \frac{\alpha}{\beta} as\ n \to \infty$, provided $\beta \neq 0$.

6.4.1.CHEBYCHEV'S THEOREM:-

As an immediate consequence of Chebychev's inequality, we have the following theorem on convergence in probability. "If $X_1, X_2, \dots\dots . X_n$ is a sequence of random variables and if mean μ_n and standard deviation σ_n of X_n exists for all n and if $\sigma_n \to 0$ as $n \to \infty$"

Proof. By Chebychev's inequality, for $\in > 0$,

$$P\{|X_n - \mu_n| \geq \epsilon\} \leq \frac{\sigma_n^2}{k^2} \to 0 \; as \; n \to \infty (\because \sigma_n \to 0 \text{ as } n \to \infty).$$

Hence $X_n - \mu_n \xrightarrow{p} 0 \; as \; n \to \infty$, provided $\sigma_n \to 0$ as $n \to \infty$.

6.5.WEAK LAW OF LARGE NUMBERS :-

Let $X_1, X_2, X_3, \dots\dots X_n \dots\dots\dots$ be a sequence of random variables and $\mu_1, \mu_2, \mu_3, \dots\dots \mu_n \dots\dots\dots$ be their respective expectations and let

$$B_n = Var(X_1 + X_2 + \cdots + X_n) < \infty$$

Then

$$P\left\{\left|\frac{X_1+X_2+\cdots+X_n}{n} - \frac{\mu_1+\mu_2+\cdots+\mu_n}{n}\right| < \epsilon\right\} \geq 1 - \eta \dots\dots\dots\dots\dots\dots (\mathbf{6.5.1})$$

For all $n > n_0$, where ϵ and η are arbitrary small positive numbers, provided

$$\lim_{n \to \infty} \frac{B_n}{n^2} \longrightarrow 0.$$

Proof. Using Chebychev's Inequality, to the random variable $\frac{(X_1+X_2+\cdots+X_n)}{n}$, we get for any $\epsilon > 0, P\left\{\left|\left(\frac{X_1+X_2+\cdots+X_n}{n}\right) - \left(E\frac{X_1+X_2+\cdots+X_n}{n}\right)\right| < \epsilon\right\} \geq 1 - \frac{B_n}{n^2 \in^2}$,

$$\left[Since \; Var\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) = \frac{1}{n^2} Var(X_1 + X_2 + \cdots + X_n) = \frac{B_n}{n^2 \in^2}\right]$$

It implies that $P\left\{\left|\left(\frac{X_1+X_2+\cdots+X_n}{n}\right) - \left(\frac{\mu_1+\mu_2+\cdots+\mu_n}{n}\right)\right| < \epsilon\right\} \geq 1 - \frac{B_n}{n^2 \in^2}$.

So far, nothing is assumed about the behaviour of B_n for indefinitely increasing values of n.Sinceϵ is arbitrary, we assume $\frac{B_n}{n^2 \in^2} \longrightarrow 0, as \; n$ becomes indefinitely large. Thus, having chosen two arbitrary small positive numbers ϵ and η, number n_0 can be found so that inequality $\frac{B_n}{n^2 \epsilon^2} < \eta$, will hold for $n > n_0$.Consequently, we shall have

$$P\left\{\left|\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) - \left(\frac{\mu_1 + \mu_2 + \cdots + \mu_n}{n}\right)\right| < \epsilon\right\} \geq 1 - \eta$$

For all $n > n_0$, (ϵ, η). This conclusion leads to the following important result, known as the (Weak) Law of Large Numbers:

"With the probability approaching unity or certain as near as we please, we may expect that the arithmetic mean of values actually assumed by n random variables will differ from the arithmetic mean of their expectations by less than any given number, however Small, provided the number of variables can be taken sufficiently large and provided the condition$\frac{B_n}{n^2} \longrightarrow 0$ as $n \longrightarrow \infty$ is fulfilled."

Remark 6.5.1.$\bar{X}_n \xrightarrow{p} \overline{\mu_n}$provided$\frac{B_n}{n^2} \longrightarrow 0$ as $n \longrightarrow \infty$.

Remark 6.5.2. For the existence of the law we assume the following conditions:

a) $E(X_i)$ exists for all i,
b) $B_n = Var(X_1 + X_2 + \cdots + X_n)$exists.
c) $\frac{B_n}{n^2} \longrightarrow 0$as $n \longrightarrow \infty$.

Remark 6.5.3. If $X_1, X_2, X_3, \ldots\ldots\ldots X_n \ldots\ldots\ldots$ are independent and identically distributed.
i.e., if $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$ for all $i = 1,2, \ldots\ldots..n$ then

$$B_n = Var(X_1 + X_2 + \cdots + X_n) = \sum_{i=1}^{n} Var\ (X_i)$$

The convariance terms vanish, since variables are independent.

$$B_n = n\sigma^2 \ldots\ldots\ldots\ldots\ldots\ldots\ldots..(\mathbf{6.5.2})$$

Hence $\lim_{n \to \infty} \frac{B_n}{n^2} = \lim_{n \to \infty} (\sigma^2/n) = 0$. Thus, the law of large number holds for the sequence $\{X_n\}$ of independent and identically distributed random variables and we get,

$P\left\{\left|\frac{X_1+X_2+\cdots+X_n}{n} - \mu\right| < \epsilon\right\} > 1 - \eta$for $n > n_0$ $i.e., P\{|\overline{X_n} - \mu| < \epsilon\} \longrightarrow 1\ as\ n \longrightarrow \infty \Rightarrow P\{|\overline{X_n} - \mu| < \epsilon\} \longrightarrow 0\ as\ n \longrightarrow \infty$.

Where $\overline{X_n}$ is the mean of the n random variables $X_1, X_2, X_3, \ldots\ldots\ldots X_n$.This result implies that $\overline{X_n}$ converges in probability to $\mu, i.e., \bar{X}_n \xrightarrow{p} \mu$.

- If $\overline{X_n}$ is the mean of n identically distributed random variables$X_1, X_2, X_3, \ldots\ldots\ldots X_n$ with if $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$, then $E(\overline{X_n}) = \mu$; and $Var(\overline{X_n}) = Var(\sum_{i=1}^{n} x_i/n) = \sigma^2/n$.

6.6.BERNOULI'S LAW OF LARGE NUMBERS:-

Jakob Bernoulli

Fig 6.5.1

Ref:

https://www.britannica.com/science/law-of-large-numbers

The law of large numbers, in statistics, the theorem that, as the number of identically distributed, randomly generated variables increases, their sample mean (average) approaches their theoretical mean.The law of large numbers was first proved by the Swiss mathematician Jakob Bernoulli in ***1713***. The law of large numbers is closely related to what is commonly called the law of averages.

Let there be n trials of an event, each trial resulting in a success or failure. If X is the number of successes in n trials with constant probability p of success for each trial, then $E(X) = np$ and $Var(X) = npq, q = 1 - p$. The variable X/n represents of successes or the relative frequency of successes, and

$E\left(\frac{X}{n}\right) = p$, and $Var\left(\frac{X}{n}\right) = \frac{1}{n^2}\, Var(X) = \frac{pq}{n}$.

Then,

$$P\left\{\left|\frac{X}{n} - p\right| < \epsilon\right\} \longrightarrow 1 \text{ as } n \longrightarrow \infty \Longrightarrow P\left\{\left|\frac{X}{n} - p\right| \geq \epsilon\right\} \text{ as } n \longrightarrow \infty$$

………………………………………… … … $(\mathbf{6.6.1})$

for any assigned $\epsilon > 0$. This implies that (X/n) converges in probability to p as $n \longrightarrow \infty$.

Proof. Using Chebychev's Inequality, to the random variable $\left(\frac{X}{n}\right)$, we get for any $\epsilon > 0, P\left\{\left|\frac{X}{n} - E\left(\frac{X}{n}\right)\right| \geq \epsilon\right\} \leq \frac{Var\,(X/n)}{\epsilon^2} = \frac{pq}{n\epsilon^2} \leq \frac{1}{4n\epsilon^2}$.

Since the maximum value of pq is at $p = q = \frac{1}{2}, i.e.,$ $\max(pq) = \frac{1}{4}, i.e., pq \leq \frac{1}{4}$. Since ϵ is arbitrary, we get

$$P\left\{\left|\frac{X}{n} - E\left(\frac{X}{n}\right)\right| \geq \epsilon\right\} \rightarrow 0 \; as \; n \longrightarrow \infty \Longrightarrow P\left\{\left|\frac{X}{n} - E\left(\frac{X}{n}\right)\right| \geq \epsilon\right\} \longrightarrow 1$$

as $n \longrightarrow \infty$.

6.6.1.MARKOV'S THEOREM:-

The weak law of large numbers holds if for some $\delta > 0$, all the mathematical expectations $E|X_i|^{1+\delta}; i = 1,2, \dots \dots \dots$ exist and are bounded. Markov theorem provides only a necessary condition for the weak law of large numbers to hold good.

6.6.2.KHINCHIN'S THEOREM:-

If $X_i's$ are identically and independently distributed random variables, the only necessary condition for the law of large numbers to hold is that $E(X_i); i = 1,2$ ....should exist.

6.7. BOREL-CANTELLI LEMMA:-

Theorem 6.7.1. Let$A_1, A_{2,}A_3, \dots \dots \dots \dots$ be a sequence of events on the probability space $(S, B, P)$and let $A = \overline{\lim} A_n$. If $\sum_{n=1}^{\infty} P(A_n) < \infty$, then $P(A) = 0$. It means that if $\sum_{n=1}^{\infty} P(A_n)$ converges then with probability one, only a finite number of $A_1, A_{2,}A_3, \dots \dots \dots$ ...can occur.

Proof. Since $A = \overline{\lim} A_n = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_n$, we have $A \subset \bigcup_{m=n}^{\infty} A_m$, for every n. Thus for each $n, P(A) \leq \sum_{m=n}^{\infty} P(A_m)$. Since $\sum_{n=1}^{\infty} P(A_n)$ is convergent (given), $\sum_{m=n}^{\infty} P(A_m)$, being the remainder term of a convergent series, tends to zero $as\ n \longrightarrow \infty$. Thus $P(A) = 0$.

Borel-Cantelli Lemma (Converse). Let $A_1, A_2, \dots$.... Be independent events on $(S, B, P)$and $A = \overline{\lim} A_n$,If $\sum_{n=1}^{\infty} P(A_n) = \infty$, then If $P(A) = 1$.

Proof. We have for any $m, n(m > n). \bigcap_{k=n}^{\infty} \overline{A_k} \subset \bigcap_{k=n}^{m} \overline{A_k} \Longrightarrow P(\bigcap_{k=n}^{\infty} \overline{A_k}) \leq P(\bigcap_{k=n}^{m} \overline{A_k}) = \prod_{k=n}^{m} P(\overline{A_k})$ (where $\overline{A_n} = S - A_n$) , because of the fact that if $(A_n, A_{n+1}, \dots \dots . A_m)$ are independent events, so are $(\overline{A_n}, \bar{A}_{n+1}, \dots \dots . \overline{A_m})$. Hence $P(\bigcap_{k=n}^{m} \overline{A_k}) = \prod_{k=n}^{m}\{1 - P(A_k)\} \leq \prod_{k=n}^{m} e^{-P(A_k)}$ $[\because 1 - x \leq e^{-x}, for\ x \geq 0] =$ $exp\{-\sum_{k=n}^{m} P(A_k)\}, \forall m$.Since $\sum_{n=1}^{\infty} P(A_k) = \infty, \sum_{k=n}^{\infty} P(A_n) \longrightarrow \infty$,as $m \longrightarrow \infty$,

therefore $P(\bigcap_{k=n}^{\infty} \overline{A_k}) = 0$ ....*(5.7.1)*

But $A = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_n$ therefore $\bar{A} = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} \bar{A}_k$(De-Morgan's Law).It implies that $P(\bar{A}) \sum_{n=1}^{\infty} P(\bigcap_{k=n}^{\infty} \overline{A_k}) = 0 \Longrightarrow P(\bar{A}) = 0$

(from ***…(5.7.1)).***

Hence $P(A) = 1 - P(\bar{A}) = 1$, as required. If $A_1, A_2, \dots$.... are independent events on (S, B, P), it follows from Theorem (5.7.1) that the probability that an infinite number of them occur is either zero $(\sum_{n=1}^{\infty} P(A_n) < \infty)$ or one $[when \sum_{n=1}^{\infty} P(A_n) = \infty)]$. This statement is a special case of so-called "Zero-one law".

Zero-one law: If $A_1, A_2, \dots$.... are independent events and if E belongs to the $\sigma -$field generated by the class $(A_n, A_{n+1,} \dots .)$ for every n, then $P(E)$ is zero or one.

6.8.PROBABILITY GENERATING FUNCTION:-

If $a_0, a_1, a_2 \dots ....$ Is a sequence of real numbers and if $A(s) = a_0 + a_1 s + a_2 s^2 + \cdots = \sum_{i=1}^{\infty} a_i s^i$converges in some interval $-s_0 < s < s_0$, when the sequence is infinite then the function $A(s)$ is known as the generating function of the sequence $\{a_i\}$. When a_i is the probability that an integral valued discrete variable X takes the values i, i.e.,$a_i = p_i = P(X = i); i = 0,1,2, \dots \dots$ with $\sum p_i = 1$, then the probability generating function, abbreviated as $p.g.f.$, of $r.v.X$ is defined as:

$P(s) = E(s^X) = \sum_{x=0}^{\infty} s^x . p_x$**..................................*(6.8.1)***

Remark 6.8.1. Obviously, we have$P(1) = \sum_x p_x = 1$. Thus a function $P(s)$ defined in ***(6.8.1)*** is a probability generating function iff$p_x \geq 0 \forall x$ and $\sum_x p_x = 1$.

Remark 6.8.2. Taking $s = e^t$in (6.8.1) we get $P(e^t) = E(e^{tX}) = M_X(t), i.e.$, from probability generating function we can obtain moment generating function on replacing s by e^t..............**(6.8.2)**

Remark 6.8.3. The joint probability generating function of two random variables X_1 and X_2 is a function of two random variables s_1 and s_2 defined by:

$$P_{X_1,X_2}(s_1, s_2) = E({s_1}^{X_1}.{s_2}^{X_2}) = \sum_{x_1}\sum_{x_2} {s_1}^{x_1}.{s_2}^{x_2} p(x_1, x_2)..(\mathbf{6.8.3})$$

Marginal probability generating function can be obtained from ***(6.8.3)*** as given below:

$$P_{X_1}(s_1) = E({s_1}^{X_1}) = P_{X_1,X_2}(s_1, 1); P_{X_2}(s_2) = E({s_2}^{2}) = P_{X_1,X_2}(1, s_2) \dots (\mathbf{6.8.4})$$

Remark 6.8.4. Two random variables variables X_1 and X_2 are independent $\Leftrightarrow P_{X_1,X_2}(s_1, s_2) = P_{X_1}(s_1).P_{X_2}(s_2) \dots \dots \dots \dots \dots (\mathbf{6.8.5})$**.**

The above concepts can be generalised to n random variables.

6.9. SOLVED EXAMPLES:-

Example 6.9.1.A symmetric die is thrown 600 times. Find the lower bound for the probability of getting 80 to 120 sixes.

Solution. Let S be total number of successes. Then$E(S) = np = 600 \times \frac{1}{6} = 100$ and $E(S) = npq = 600 \times \frac{1}{6} \times \frac{5}{6} = \frac{500}{6}$. Using Chebyshev's inequality, we get,

$P\{|X - E(S)| < k\sigma\} \geq 1 - \frac{1}{k^2} \Rightarrow P\left\{|X - 100| < k\sqrt{\frac{500}{6}}\right\} \geq 1 - \frac{1}{k^2}$.

Therefore $P\left\{100 - k\sqrt{\frac{500}{6}} < S < 100 + k\sqrt{\frac{500}{6}}\right\} \geq 1 - \frac{1}{k^2}$

Taking $k = \frac{20}{\sqrt{500/6}}$

$$P\{80 \leq S \leq 120\} \geq 1 - \frac{1}{400 \times (6/500)} = \frac{19}{24}$$

Example 6.9.2. UseChebyshev's inequality to determine how many times a fair coin must be tossed in order that the probability will be at least 0.90 that the ratio of the observed number of heads to the number of tosses will lie between 0.4 and 0.6.

Solution. By Bernoulli's Law of Large Numbers, we get for any$\epsilon > 0$,

$$P\left\{\left|\frac{X}{n} - p\right| < \epsilon\right\} \longrightarrow 1 \; as \; n \longrightarrow \infty \Rightarrow P\left\{\left|\frac{X}{n} - p\right| \geq \epsilon\right\} \; as \; n \longrightarrow \infty$$

Since $p = 0.5(as\ the\ coin\ is\ unbiased)$ and we want the proportion of successes X/n to lie between 0.4 and 0.6,we have,$\left|\frac{X}{n} - p\right| \leq 0.1$. Thus choosing $\epsilon = 0.1$, we have $P\left\{\left|\frac{X}{n} - p\right| < 0.1\right\} \geq 1 - \frac{1}{4n(0.1)^2} = 1 - \frac{1}{0.04}$. Since we want this probability to be 0.9, we fix $1 - \frac{1}{0.04n} = 0.90$ it implies that $0.10 = \frac{1}{0.04n} \Rightarrow n = \frac{1}{0.10 \times 0.04} = 250$.

Hence the required number of tosses is 250.

Example 6.9.3. Two unbiased dice are thrown. If X is the sum of the numbers showing up, prove that$P\{|X - 7| < k\sigma\} \geq 3 \leq \frac{35}{54}$.Compare this with the actual probability.

Solution. The probability distribution of the random variable$X$ (the sum of the numbers on the two dice) is as given in the following table:

X	**Favourable cases (distinct)**	**Probability (p)**
2	(1,1)	$1/36$
3	(1,2),(2,1)	$2/36$
4	(1,3),(3,1)(2,2)	$3/36$
5	(1,4),(4,1),(2,3),(3,2)	$4/36$
6	(1,5),(5,1),(2,4),(4,2),(3,3)	$5/36$
7	(1,6),(6,1),(2,5),(5,2),(3,4),(4,3)	$6/36$
8	(2,6),(6,2),(3,5),(5,3),(4,4)	$5/36$
9	(3,6),(6,3),(4,5),(5,4)	$4/36$
10	(4,6),(6,4),(5,5)	$3/36$
11	(5,6),(6,5)	$2/36$
12	(6,6)	$1/36$

$$E(X) = \sum_x p.x$$

$= \frac{1}{36}(2 + 6 + 12 + 20 + 30 + 42 + 40 + 36 + 30 + 22 + 12) = 7$

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

$$= \frac{1}{36}(4 + 18 + 48 + 100 + 180 + 294 + 320 + 324 + 300 + 144) = \frac{1}{36}(1974) = \frac{329}{6}$$

Therefore $Var\ (X) = E(X^2) - \{E(X)\}^2 = \frac{329}{6} - (7)^2 = \frac{35}{6}$

Using Chebyshev's inequality, we get

$P\{|X - \mu| \geq k\} \leq \frac{Var\ X}{k^2} \Longrightarrow P\{|X - 7| \geq 3\} \leq \frac{35/6}{9} = \frac{35}{54}$.(Taking $k = 3$)

Actual Probability:

$P\{|X - 7| \geq 3\} = 1 - P\{|X - 7| < 3\}$

$= 1 - P(4 < X < 10)$.

$= 1 - [P(X = 5) + P(X = 6) + P(X = 7) + P(X = 8) + P(X = 9)]$

$$= 1 - \frac{1}{36}[4 + 5 + 6 + 5 + 4] = 1 - \frac{24}{36} = \frac{1}{3}$$

Example 6.9.4: Let $X_1, X_2, \dots \dots \dots \dots \dots, X_n$ independent and identical distributed variables with mean $\mu$and variance $\sigma^2$and as $n \longrightarrow \infty, (x_1^2 + x_2^2 + \cdots x_n^2)/n \overset{p}{\to} c$,for some constant c, for some constant c; $(0 \le c \le \infty)$. Find c.

Solution. $E(X_i) = \mu, Var\ (X_i) = \sigma^2; i = 1,2, \dots \dots \dots n.$

Therefore $E(X_i^2) = Var(X_i) + \{E(X_i)\}^2 = \sigma^2 + \mu^2(finite); i = 1,2, \dots \dots \dots n.$

Since $E(X_i^2)$ is finite; by khinchine's Theorem weak law of large number holds for the sequence of independent and identical distributed variables for the sequence X_i^2 of independent and identical distributed random variables so that,

$$(x_1^2 + x_2^2 + \cdots x_n^2)/n \overset{p}{\to} E(X_i^2)\ as\ n \longrightarrow \infty .$$

This implies that

$$(x_1^2 + x_2^2 + \cdots x_n^2)/n \overset{p}{\to} \sigma^2 + \mu^2 = c\ as\ n \longrightarrow \infty$$

Hence $c = \sigma^2 + \mu^2$.

Example 6.9.5:A bag contains one black ball and m white balls. Aball is drawn at random. If a white ball is drawn, it is returned to the bag together with an additional white ball. If the black ball is drawn, it alone is returned to the bag.

Let A_n denote the event that the black ball is not drawn in the first n trials. Discuss the converse to Borel-Cantelli Lemma with reference to events $A_n, n = 1,2,3 \dots \dots \dots ..$

Solution:

$A_n =$ The event that black ball is not drawn in the first n trials ....***(6.9.4.1)***

= The event that each of the first n trials resulted in the draw of a white ball.

$\Rightarrow \mathrm{P(A_n) = P(E_1 \cap E_2 \cap \dots \dots \cap E_n)}$,where$\mathrm{E_i}$ is the event of drawing a white ball

in the ith trial.Therefore,

$$\mathrm{P(A_n)} = \mathrm{P(E_1)} P(E_2/E_1). P(E_3/E_1 \cap E_2).. P(E_n/E_1 \cap E_2..\cap E_{n-1})$$

$$= \frac{m}{m+1} \times \frac{m+1}{m+2} \times \dots\dots\dots\dots \times \frac{m+n-1}{m+n}$$
$$= \frac{m}{m+n} \dots\dots . (\mathbf{6.9.4.2})$$

(Since if first ball drawn is white (W) it is returned together with an additional white ball, i.e., for the second draw the box contains $1B, (m+1)W$ balls and

$$P(E_2/E_1) = \frac{m+1}{m+2}, and\ so\ on.$$

Therefore,

$\sum_{n=1}^{\infty} \mathrm{P}(\mathrm{A_n}) = \sum_{n=1}^{\infty} \frac{m}{m+n} = m \sum_{n=1}^{\infty} \frac{1}{m+n} = m \left(\frac{1}{m+n} + \frac{1}{m+2} + \frac{1}{m+3} + \cdots \right)$

$$\sum_{n=1}^{\infty} \mathrm{P}(\mathrm{A_n}) = = m \left\{ \sum_{n=1}^{\infty} \frac{1}{n} - \left(\sum_{n=1}^{m} \frac{1}{n} \right) \right\} \dots\dots . (\mathbf{6.9.4.3})$$

Since $\sum_{n=1}^{\infty} \frac{1}{n}$ is divergent by p –series test. Hence $\sum_{n=1}^{\infty} \mathrm{P}(\mathrm{A_n}) = \infty$. From the definition of $\mathrm{A_n}$ in **…(6.9.4.1)** it is obvious that $\mathrm{A_n} \downarrow$. Therefore $A = \overline{\lim_n} = \lim_n sup A_n = \emptyset \Longrightarrow P(A) = P(\emptyset) = 0$. This result is inconsistent with the converse of Borel-Cantelli Lemma, the reason being that the events $\mathrm{A_n} (n = 1,2, \dots\dots)$considered here are not independent.Since $\mathrm{P}(\mathrm{A_i} \cap A_j) = \mathrm{P}(\mathrm{A_i}) = \frac{m}{m+i} \neq P(A_i)P(A_j)$. Since for $(i > j), A_i \subset A_j\ as\ A_n \downarrow$.[from $(\mathbf{6.9.4.2})$**].**

Example 6.9.6. Find the probability generating function of :

a) $P(X \leq n)$,
b) $P(X < n)$,
c) $P(X = 2n)$,

Solution. a) Let X be an interval valued random variable with the probability distribution :

$P(X = n) = p_n$and$P(X \leq n) = q_n$.So that $q_n = p_0 + p_{1+\dots .}p_n; n = 0,1,2 \dots ..$

Therefore $q_n - q_{n-1} = p_n, n \geq 1$. It implies $\sum_{n=1}^{\infty} q_n s^n - \sum_{n=1}^{\infty} q_{n-1} s^n = \sum_{n=1}^{\infty} p_n s^n$. It implies that $Q(s) - q_0 - sQ(s) = P(s) - p_0$. Hence,

$$Q(s) = \frac{P(s) + q_0 - p_0}{1-s} = \frac{P(s)}{1-s} [\because q_0 = p_0]$$

b)Let $P(X < n) = q_n .= p_0 + p_{1+\cdots}.p_{n-1,}; q_n - q_{n-1} = p_{n-1,}, n \geq 2$.It implies $\sum_{n=2}^{\infty} q_n s^n - \sum_{n=2}^{\infty} q_{n-1} s^n = \sum_{n=2}^{\infty} p_{n-1} s^n = s \sum_{n=1}^{\infty} q_n s^n$. It implies that $Q(s) - q_1 s - sQ(s) = sP(s) - sp_0 . [\because q_0 = 0] \Rightarrow Q(s)[1-s] = sP(s) - sp_0 + q_1 s\ [\because q_0 = p_0]$

c)Let $P(X = 2n) .= p_{2n}$.Its $p.g.f.Q(s)$ is given by:

$$Q(s) = \sum_{n=0}^{\infty} p_{2n} s^n = p_0 + p_2 s + p_4 s^2 + \cdots$$

$$2Q(s) = 2p_0 + 2p_2 s + 2p_4 s^2 + \cdots = \left(p_0 + p_1 s^{1/2} + p_2 s + p_3 s^{3/2} + p_4 s^2 + \cdots\right) + \left(p_0 - p_1 s^{1/2} + p_2 s - p_3 s^{3/2} + p_4 s^2 + \cdots\right)$$

$$2Q(s) = \sum_{k=0}^{\infty} p_k \left(s^{1/2}\right)^k + \sum_{k=0}^{\infty} p_k \left(-s^{1/2}\right)^k$$

$$= P\left(s^{1/2}\right) + P\left(-s^{1/2}\right)$$

$$Q(s) = \frac{P\left(s^{1/2}\right) + P\left(-s^{1/2}\right)}{2}$$

CHECK YOUR PROGRESS

Problem 1. Using Chebyshev's inequality, calculate the percentage of observations that would fall outside 3 standard deviations of the mean.

a) 11%
b) 89%
c) 90%
d) 72%

Problem 2.A class of second graders has a mean height of five feet with a standard deviation of one inch. At least what percent of the class must be between 4'10"and 5'2"?

a) 75%
b) 80%
c) 90%
d) 95%

Problem3. The law of large numbers shows a relationship between the theoretical probability and the

a) Sample size
b) exponential probability
c) experimental probability
d) rational probability

Problem 4. The practical result of the central limit theorem is that

a) researchers must take a large number of samples before inferences about the population can be made.
b) The researcher must know the shape of the population distribution before inferences about the population can be made.
c) Small-sized samples should not be used in research.
d) The concept of the sampling distribution is unimportant to researchers.
e) none of the above

Problem4. The formula Z = (X - μ)/σ where μ is the hypothesized or expected value of the mean,

a) is the formula for a confidence interval.
b) The way of making a linear transformation of any normal variable into a standard normal variable.
c) The formula for calculating the standard error of the mean.
d) The computation for estimating the value of the central limit theorem.

Problem6. For Chebyshev's inequality, the k must be an integer. True\False. True\False.

Problem7. The Chebyshev's inequality also tells us

$P\{|X-\mu| \geq k\sigma\} \leq \frac{1}{k^2}$. True\False.

Problem8. Chebyshev's inequality can help us estimate$P(\mu-\sigma \leq X \leq \mu+\sigma)$. True\False.

Problem9. We can use Chebyshev's inequality to prove the Law of Large Numbers. True\False.

Problem10. A coin is weighted so that its probability of landing on heads is 20%. Suppose the coin is flipped 20 times. The bound for the probability it lands on heads at least 16 times is 1.38×10^{-8}.True\False.

6.10. SUMMARY:-

In this unit first we explained about Chebychev's Inequality. Chebyshev's inequality is a probabilistic inequality then convergence in Probability is explained here. The idea of Weak Law of Large Numbers (W.L.L.N) is also discussed here. After that Some other concept of Large number is also explained. After that concept of Bernouli's Law of Large Numbers is defined then Summarize the Khinchin's Theorem, Borel-Cantlli Lemma (Zero-One Law) and explained the concept of Probability Generating Function (p.g.f).

6.11.GLOSSARY:-

i. Random variable
ii Mean
iii Variance
iv Moments
v Mathematical expectation
vi Convergence
vii Probability Space
viii Independent events.
ix Moment generating function.

6.12. REFERENCE BOOKS:-

1. S. C. Gupta and V. K. Kapoor, (2020), *Fundamentals of mathematical statistics*, Sultan Chand & Sons.
2. Seymour Lipschutz and John J. Schiller, (2017), *Schaum's Outline: Introduction to Probability and Statistics*, McGraw Hill Professional.
3. J. S. Milton and J. C. Arnold , (2003*), Introduction to Probability and Statistics (4th Edition),* Tata McGraw-Hill.
4. https://www.wikipedia.org.

6.13.SUGGESTED READINGS:-

1. A.M. Goon,(1998), *Fundamental of Statistics (7th Edition),* 1998.
2. R.V. Hogg and A.T. Craig, (2002), *Introduction to Mathematical Statistics,* MacMacMillan, 2002.
3. Jim Pitman, (1993), *Probability,* Springer-Verlag.
4. https://archive.nptel.ac.in/courses/111/105/111105090.

6.14.TERMINAL QUESTIONS:-

i If you wish to estimate the proportion of engineers and scientists who have studied probability theory and you wish your estimate to be correct within 2% with probability 0.95 or more, how large a sample would you take (a) if you have no idea what the true proportion is, (b) if you are confident that the true proportion is less than 0.2?

ii For geometric distribution $p(x) = 2^{-x}; x = 1,2,3, \dots \dots$prove that Chebychev's inequality gives $P\{|x-2| \leq 2\} > \frac{1}{2}$, while the actual probability is $\frac{15}{16}$.

iii State and Prove Chebychev's inequality?..

iv Let $f(x) = \frac{5}{x^6}$ for $x \geq 1$and 0otherwise. What bound does Chebyshev's inequality give for the probability$P \geq 2.5$?For what value of a can we say$P \geq a \leq 15\%$?

6.15.ANSWERS:-

Answer of Check your progress Questions:-

i. A
ii. 75%
iii. Experimental Probability
iv. e
v. b
vi. False
vii. True
viii. True
ix. False
x. True

Answer of Terminal Questions:-

i. $(a)\ n = 12{,}500.\ (b) 8{,}000.$
iv $1/15$, $25/12$

BLOCK III
PROBABILITY DISTRIBUTIONS

UNIT:-7 DISCRETE PROBABILITY DISTRIBUTIONS

CONTENTS:

7.1 .INTRODUCTION:-

In previous unit we have discussed about Chebychev's Inequality, Convergence in Probability,Weak law of large numbers (W.L.L.N.), Bernoulli's Law of Large Number, Borel-Cantelli Lemma (Zero-One Law) and Probability Generating Function (p.g.f). In this unit we explained about Discrete uniform Distribution, Bernoulli Distribution, Binomial Distribution, Possion Distribution, Negative Binomial Distribution, Geometric Distribution and Hypergeometric Distribution. In probability theory and statistics, a probability distribution is the mathematical function that gives the probabilities of occurrence of different possible outcomes for an experiment.

7.2.OBJECTIVES:-

After studying this unit learner will be able to:

i Write down expressions for different discrete distribution.
ii Calculate the mean and variance of the different discrete distributions.
iii Show that Poisson distribution is a limiting case of binomial distribution.

7.3.DISCRETE UNIFORM DISTRIBUTION:-

A random variable $(r.v.)\ X$ is said to have a discrete uniform distribution over the range $[1, n]$ if its $p.m.f.$ is expressed as follows:

$$P(X = x) = \begin{cases} \frac{1}{n}\ for\ x = 1,2, \dots n \\ 0, otherwise \end{cases} \text{..........................}\textbf{(7.3.1)}$$

Here n is known as the parameter of the distribution and lies in the set of all positive integers. Equation (7.3.1) is also called a discrete rectangular distribution.

- A simple example of the discrete uniform distribution is throwing a fair die. The possible values are 1, 2, 3, 4, 5, 6, and each time the die is thrown the probability of a given score is 1/6. If two dice are thrown and their values added, the resulting distribution is no longer uniform because not all sums have equal probability.

Moments. $E(X) = \frac{1}{n}\sum_{i=1}^{n} i = \frac{n+1}{2}, E(X^2) = \frac{1}{n}\sum_{i=1}^{n} i^2 = \frac{(n+1)(2n+1)}{6}$

$$V(X) = E(X^2) - [E(X)]^2 = \frac{(n+1)(n-1)}{12}$$

The moment generating function of X is :

$$M_X(t) = E(e^{tX}) = \frac{1}{n}\sum_{x=1}^{n} e^{tx} = \frac{e^t(1 - e^{nt})}{n(1 - e^t)}$$

7.4.BEROULLI DISTRIBUTION:-

A random variable $(r.v.)\ X$ is said to have a Bernoulli distribution with parameter $p$if its $p.m.f.$ is given by:

$$P(X = x) = \begin{cases} p^x(1 - p)^{1-x}\ \ for\ x = 1,2, \dots n \\ 0,\ otherwise \end{cases} \text{.........}\textbf{\textit{(7.3.2)}}$$

The parameter p satisfies $0 \le p \le 1$. Often $(1 - p)$ is denoted q.

A random experiment whose outcomes are two types, success S and failure F, occurring with probabilities p and failure q respectively, is called a Bernoulli trial. If for this experiment, a random variable X is defined such that it takes values 1 when S occurs and 0 if F occurs, then X follows a Bernoulli distribution.

7.5.BINOMIAL DISTRIBUTION:-

Binomial distribution was discovered by ***James Bernouli (1654-1705)*** in the year ***1700*** and was first published in ***1713.***

Let a random experiment be performed repeatedly, each repetition being called a trial and let the occurrence of an event in a trial be called a success and its non-occurrence a failure. Consider a set of n independent Bernoullian trials (n being finite) in which the probability $'p'$ of success in any trial is constant for each trial, then $q = 1 - p$, is the probability of failure in any trial.

The probability of x successes and consequently $(n - x)$ failure in n independent trials, in a specified order (say) $SSFSFFFS \ldots\ldots FSF$ (where S represents success and F represent failure) is given by the compound probability theorem by the expression:

$$P(SSFSFFFS \ldots\ldots FSF)$$
$$= P(S)P(S)P(F)P(S)P(F)P(F)P(S) \times P(F)P(S)P(F)$$
$$= p.p.q.p.q.q.q.p \ldots.q.p.q$$
$$= \underset{\{x\ factors\}}{p.p.p.p \ldots\ldots p} \,.\, \underset{\{(n-x)\ factors\}}{q.q.q \ldots.q} = p^x q^{n-x}$$

But x successes in n trials can occur in $\binom{n}{x}$ ways and the probability for each of these ways is same, viz., $p^x q^{n-x}$. Hence the probability of x successes in n trials in any order is given by the addition theorem of probability by the expression $\binom{n}{x}\ p^x q^{n-x}$.

The probability distribution of the number of successes, so obtained is called the Binomial probability distribution.

The probabilities of $0,1,2, \ldots\ldots, n$ successes, viz., $q^n, \binom{n}{1}q^n p, \binom{n}{1}q^n\ p^2, \ldots., \ p^n$, are the Binomial expansion of $(q + p)^n$.

> *A random variable X is said to follow binomial distribution if it assumes only non-negative values and its probability mass function is given by:*

$$P(X = x) = p(x) = \begin{cases} \binom{n}{x}\ p^x q^{n-x}; x = 0,1,2 \ldots\ldots\ldots n; q = 1 - p \\ 0, \quad Otherwise \end{cases} \ldots\ldots\ldots\ldots\ldots\ldots \textbf{\textit{(7.5.3)}}$$

- n and p in the distribution are known as the parameters of the distribution.
- Degree of Binomial distribution is sometimes $'n'$.

- The value of $X = 0, 1, 2, \dots\dots.., n$.
- Any random variable which follows binomial distribution is known as binomial variate.
- $X \backsim B(n, p)$ to random variable X follows binomial distribution with parameters n and p.
- The probability $p(x)$ in ***(7.5.3)*** is also sometimes denoted by $b(x, n, p)$.
- The assignment of probabilities in (7.5.3) is permissible because
$$\sum_{x=0}^{n} p(x) = \sum_{x=0}^{n} \binom{n}{x} p^x\, q^{n-x} = (q + p)^n = 1$$
- Let us suppose that n trials constitute an experiment. Then, if this experiment is repeated N times, the frequency function of the Binomial distribution is given by:
$$f(x) = Np(x) = N \sum_{x=0}^{n} \binom{n}{x} p^x\, q^{n-x}; x = 0,1,2, \dots\dots\dots, n \quad \dots.. \quad (\mathbf{7.5.4})$$
and the expected frequencies of $0,1,2, \dots\dots\dots, n$ successes are the successive terms of the binomial expansion, $N(q + p)^n, q + p = 1$.
- In binomial distribution under the following experimental conditions:
    i. Each trial results in two exhaustive and mutually disjoint outcomes, termed as success and failure.
    ii. The number of trials $'n'$ is finite.
    iii. The trials are independent of each other.
    iv. The probability of success $'p'$ is constant for each trial.
- The trials satisfying the condition $(i), (iii)$, and (iv) are also called *Bernoulli trials.*
- The problems relating to tossing of a coin or throwing of dice or drawing cards from a pack of cards with replacement lead to binomial probability distribution.
- Binomial distribution is important not only because of it's wide applicability, but because it gives rise to many other probability distributions.
- Tables for $p(x)$ are available for various values of $'n'$and $'p'$.

Moments of Binomial Distribution:

The four moments about origin of binomial distribution are obtained as follows:

$$\mu_1' = np,\ \mu_2' = n(n-1)p^2 + np,$$
$$\mu_3' = n(n-1)(n-2)p^3 + 3n(n-1)p^2 + np$$
$$\mu_4' = n(n-1)(n-2)(n-3)p^4 + 6n(n-1)(n-2)p^3 + 7n(n-1)p^2 + np$$

$\mu_2 = npq$, $\mu_3 = npq(1-2p)$, $\mu_4 = npq\{1 + 3(n-2)pq\}$

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{(1-2p)^2}{npq},$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = 3 + \frac{1-6pq}{npq}$$

$\gamma_1 = +\sqrt{\beta_1} = \frac{q-p}{\sqrt{npq}} = \frac{1-2p}{\sqrt{npq}}$ and $,\gamma_2 = \beta_2 - 3 = \frac{1-6p}{npq}$

Mean = $\mu = np$

Variance = $npq = \sigma^2$

- Variance is less than mean.

Recurrence Relation for the moments of Binomial Distribution (Renovsky Formula).

$$\mu_{r+1} = pq\left(nr\mu_{r-1} + \frac{d\mu_r}{dp}\right) \quad \textbf{....................(7.5.5)}$$

$\mu_2 = npq = \sigma^2$

$\mu_3 = npq(q-p)$

$\mu_4 = npq[1 + 3pq(n-2)]$

Mode of Binomial Distribution.

Case I. When $(n+1)p$ is not an integer.

$p(x)$ is maximum at $x = m$.

Case II. When $(n+1)p$ is an integer.

In this case the binomial distribution is bimodal and the two modal values are m and $m-1$.

Moment Generating Function of Binomial Distribution.

$$M_X(t) = (q + pe^t)^n \ldots\ldots\ldots\ldots\ldots\ldots (\mathbf{7.5.6})$$

Additive Property of Binomial Distribution.

- The sum of two independent binomial variates is not a binomial variate.
- The binomial distribution possesses the additive or reproductive property if $p_1 = p_2$.
- If $X_i (i = 1,2,\ldots..k)$ are independent binomial variates with parameters $(n_i, p), (i = 1,2,\ldots..k)$, then their sum $\sum_{i=1}^{k} X_i \backsim B\left(\sum_{i=1}^{k} n_i, p\right)$.

Chracteristic Function of Binomial Distribution.

$$\emptyset_X(t) = \left(q + pe^{it}\right)^n \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (\mathbf{7.5.7})$$

Probability Generating Function of Binomial Distritution.

$$P(s) = (ps + q)^n \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (\mathbf{7.5.8})$$

Cumulants of the Binomial Distribution.

Cumulant generating function is given by:

$$K_X(t) = \log M_X(t)$$
$$= n\left[p\left(t+\frac{t^2}{2!}+\frac{t^3}{3!}+\frac{t^4}{4!}+.\right) - \frac{p^2}{2}\left(t+\frac{t^2}{2!}+\frac{t^3}{3!}+\frac{t^4}{4!}+\cdots\right)^2 + \frac{p^3}{3}\left(t+\frac{t^2}{2!}+\frac{t^3}{3!}+\frac{t^4}{4!}+\cdots\right)^3+\cdots\right]$$

$$k_1 = np, k_2 = npq, k_3 = npq(q-p), k_4 = npq(1-6pq)$$

7.6.POISSON DISTRIBUTION:-

Poisson distribution was discovered by French mathematician and physicist ***Simeon Denis Poisson (1781-1840)*** who published it in ***1837***. Poisson distribution is a limiting case of the binomial distribution under the following conditions:

i n, the number of trials is indefinitely large, i.e., $n \longrightarrow \infty$.

ii p,the constant probability of success for each trial is indefinitely small, i.e., $p \longrightarrow 0$.

iii $np = \lambda, (say)$is finite.

Thus $p = \lambda/n, q = 1 - \lambda/n$, where λ is a positive real number.

The probability of x successes in a series of n independent trials is:

$$b(x;n,p) = \binom{n}{x} p^x q^{n-x}; x = 0,1,2 \dots n; q = 1-p \quad \dots (\mathbf{7.6.1})$$

We want the limiting form of ***(7.6.1)*** under the above conditions. Hence

$$\lim_{n\to\infty} b(x;n,p) = \lim_{n\to\infty} \frac{n!}{x!(n-x)!}\left(\frac{\lambda}{n}\right)^x . \left(1-\frac{\lambda}{n}\right)^{n-x}.$$

Using Stirling's approximation for $n!$as $n \to \infty$, $viz.$,

$\lim_{n\to\infty} n! \approx \sqrt{2\pi} e^{-n} n^{n+(1/2)}$,we get

$$\lim_{n\to\infty} b(x;n,p) = \lim_{n\to\infty}\left\{\frac{\sqrt{2\pi}e^{-n}n^{n+(1/2)}}{x!\sqrt{2\pi}e^{-(n-x)}(n-x)^{n-x+(1/2)}}\right\}\left(\frac{\lambda}{n}\right)^x . \left(1-\frac{\lambda}{n}\right)^{n-x}$$

$$= \frac{\lambda^x}{e^x . x!} . \lim_{n\to\infty}\frac{n^{n-x+(1/2)}}{(n-x)^{n-x+(1/2)}} . \left(1-\frac{\lambda}{n}\right)^{n-x}$$

$$= \frac{\lambda^x}{e^x . x!} . \frac{\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n . \lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{-x}}{\lim_{n\to\infty}\left(1-\frac{x}{n}\right)^n \lim_{n\to\infty}\left(1-\frac{x}{n}\right)^{-x+(1/2)}}.$$

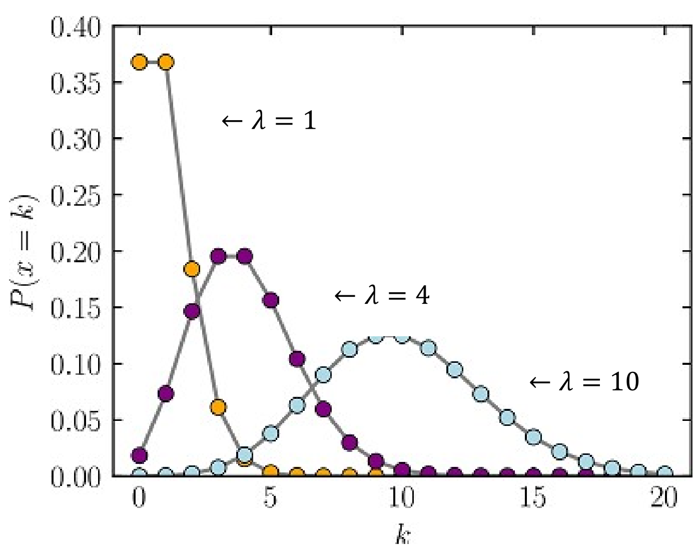
But $\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda}$, $\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^\alpha = 1, \alpha$ is not a function of n. ..**(7.6.2).** Therefore, $\lim_{n\to\infty} b(x;n,p) = \frac{\lambda^x}{e^x.x!} . \frac{e^{-\lambda}.1}{e^{-x}.1} = \frac{e^{-\lambda}\lambda^x}{x!}; x = 0,1,2, \dots \dots . \infty$ **[Using (7.6.2)].**

Which is the required probability function of the Poisson distribution 'λ' is known as the parameter of Poisson distribution.

A random variable X is said to follow a Poisson distribution if it assumes only non-negative values and it's probability mass function is given by:

$$p(x,\lambda) = P(X = x) = \begin{cases} \frac{e^{-\lambda}\lambda^x}{x!}; x = 0,1,2, \dots \dots .; \lambda > 0 \\ 0 \ otherwise \end{cases}$$

Here λ is known as the parameter of the distribution. We shall use the notation $X \sim P(\lambda)$, to denote that X is a Possion variate with parameter λ.

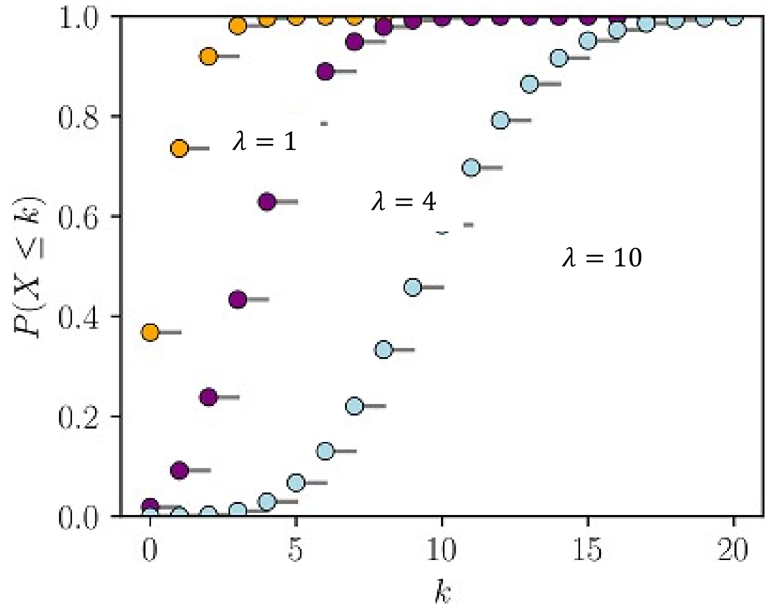


Ref 7.6.1

Probability Mass Function = $\frac{e^{-\lambda}\lambda^k}{k!}$.

https://en.wikipedia.org/wiki/Poisson_distribution#/media/File:Poisson_pmf.svg

In above figure the horizontal axis is the index k, the number of occurrences. $\lambda$is the expected rate of occurrences. The vertical axis is the probability of k occurrences given λ. The function is defined only at integer values of k; the connecting lines are only guides for the eye.



Cumulative Distribution Function $= e^{-\lambda} \sum_{j=0}^{\lfloor k \rfloor} \frac{\lambda^j}{j!}$

Ref 7.6.2
https://en.wikipedia.org/wiki/Poisson_distribution#/media/File:Poisson_pmf.svg

In above figure the horizontal axis is the index$k$, the number of occurrences. The CDF is discontinuous at the integers of k and flat everywhere else because a variable that is Poisson distributed takes on only integer values.

- $\sum_{x=0}^{\infty} P(X = x) = 1.$
- Corresponding distribution function is
$$F(x) = P(X \leq x) = \sum_{r=0}^{x} P(r) = e^{-\lambda} \sum_{r=0}^{x} \frac{\lambda^r}{r!}; x = 0,1,2, \dots \dots$$
- Poisson distribution occurs when there are events which do not occour as outcomes of a definite number of trials (unlike that in binomial distribution) of an experiment but which occur at random points of time and space wherein our interest lies only in the number of occurrence of the event, not in it's non-occurrence.

Examples: Calls per Hour at a Call Center: Call centers use the Poisson distribution to model the number of expected calls per hour that they'll receive so they know how many call center reps to keep on staff.

Number of Arrivals at a Restaurant: Restaurants use the Poisson distribution to model the number of expected customers that will arrive at the restaurant per day.

Moments of Poisson distribution:
The four moments about origin of binomial distribution are obtained as follows:

$$\mu_1' = \lambda,\ \mu_2' = \lambda^2 + \lambda,$$
$$\mu_3' = \lambda^3 + 3\lambda^2 + \lambda$$
$$\mu_4' = \lambda^4 + 6\lambda^3 + 7\lambda^2 + \lambda \mu_2 = \lambda,\ \mu_3 = \lambda, \mu_4 = 3\lambda^2 + \lambda$$
$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{\lambda^2}{\lambda^3} = \frac{1}{\lambda}$$
$$\beta_2 = \frac{\mu_4}{\mu_2^2} = 3 + \frac{1}{\lambda}$$
$$\gamma_1 = +\sqrt{\beta_1} = \frac{1}{\sqrt{\lambda}} \text{and} \quad , \gamma_2 = \beta_2 - 3 = \frac{1}{\lambda} \dots\dots\dots\dots\dots\dots\dots \mathbf{(7.6.3)}$$

Recurrence Relation for the moments of Poisson Distribution.

$$\mu_{r+1} = r\lambda\mu_{r-1} + \lambda\frac{d\mu_r}{dp} \dots\dots\dots\dots\dots\dots\dots\dots \mathit{(7.6.4)}$$
$$\mu_2 = \lambda$$
$$\mu_3 = \lambda$$
$$\mu_4 = 3\lambda^2 + \lambda$$

Mode of Poisson Distribution.

Case I. λ is not an integer.
$p(S)$is maximum value. Where S is the integral part ofλ.
Case II. $\lambda = k$isan integer
In this case we have two modal values, *viz.,* $p(k-1)$ and $p(k)$. The modes are $(\lambda - 1)$ and λ.

Moment Generating Function of Poisson Distribution.

$$M_X(t) = e^{\lambda(e^t - 1)} \dots\dots\dots\dots\dots\dots (\mathbf{7.6.5})$$

Additive Property of Poisson Distribution

The sum of two independent Poisson variatesis Poissonvariate. The converse is

Chracteristic Function of Binomial Distribution

$$\emptyset_X(t) = e^{\lambda(e^{it} - 1)} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (\mathbf{7.6.6})$$

Cumulants of the Poisson Distribution

AllCumulants of the Poisson Distribution are equal to λ.

Probability Generating Function of Binomial Distritution.

$$P(s) = e^{\lambda(s-1)} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (\mathbf{7.6.7})$$

Similarities & Differences between Binomial and Poisson distribution

Similarities:

- Both distributions can be used to model the number of occurrences of some event.
- In both distributions, events are assumed to be independent.

Difference:

- In a Binomial distribution, there is a fixed number of trials (e.g. flip a coin 3 times)
- In a Poisson distribution, there could be any number of events that occur during a certain time interval (e.g. how many customers will arrive at a store in a given hour?)

7.7 :-NEGATIVE BINOMIAL DISTRIBUTION

A random variable X is said to follow negative binomial distribution with parameter r and p if its probability mass function is given by:

$$P(X = x) = p(x) = \begin{cases} \binom{x+r-1}{r-1} \ p^x q^x; x = 0,1,2 \dots n; q = 1 - p \\ \quad 0, \quad Otherwise \end{cases} \quad \dots\dots\dots\dots(7.7.1)$$

If $p = \frac{1}{Q}$and $q = \frac{P}{Q}$, therefore $Q - P = 1$.

$$p(x) = \begin{cases} \binom{-r}{x} Q^{-r} \left(-\frac{P}{Q}\right)^x ; x = 0,1,2 \dots n; \\ \quad 0, \quad Otherwise \end{cases} \quad \dots\dots\dots\dots\dots(7.7.2)$$

Moments of Binomial Distribution:

The four moments about origin of binomial distribution are obtained as follows:

$\mu_1' = rp,\ \mu_2' = rp + r(r+1)P^2,$

$\mu_2 = rPQ,$

Mean < Variance

Moment Generating Function of Binomial Distribution.

$$M_X(t) = (Q - Pe^t)^{-r} \dots\dots\dots\dots\dots\dots (\mathbf{7.7.3})$$

Cumulants of the Binomial Distribution.

Cumulant generating function is given by:

$$K_X(t) = \log M_X(t)$$
$$= -rlog\left[1 - P\left(t + \frac{t^2}{2!} + \frac{t^3}{3!} + \frac{t^4}{4!} +.\right)\right]$$
$$k_1 = rP, k_2 = rPQ, k_3 = rPQ[Q+P], k_4 = rPQ[1 + 3PQ(r+2)] \quad \ldots\ldots\ldots\ldots\ldots\ldots (7.7.4)$$

Probability Generating Function of Binomial Distritution.

$$P(s) = [p/(1-qs)]^r \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (\mathbf{7.7.5})$$

- **Negative binomial distribution tends to Poisson distribution as** $P \longrightarrow 0, r \longrightarrow \infty$ such that $rP = \lambda(finite)$.

7.8. GEOMETRIC DISTRIBUTION:-

> *A random variable X is said to follow geometric distribution if it assumes only non-negative values and if its probability mass function is given by:*

$$P(X = x) = p(x) = \begin{cases} q^x x; x = 0,1,2 \ldots n; 0 < p \leq; q = 1-p \\ \quad 0, \quad Otherwise \end{cases} \ldots\ldots\ldots\ldots\ldots (7.8.1)$$

Moments of Binomial Distribution:

The four moments about origin of binomial distribution are obtained as follows:

$$\mu_1' = \frac{q}{p},\ Var(X) = \mu_2 = \frac{q}{p^2},$$

Moment Generating Function of Binomial Distribution.

$$M_X(t) = p(1-qe^t)^{-r} \ldots\ldots\ldots\ldots\ldots\ldots (\mathbf{7.8.2})$$

Probability Generating Function of Binomial Distritution.

$$P_X(s) = [p/(1-qs)] \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (\mathbf{7.8.3})$$

7.9 :-HYPERGEOMETRIC DISTRIBUTION

A discrete random variable X is said to follow hyper geometric distribution with parameter N, M and n if it assumes only non-negative values and its probability mass function is given by:

$$P(X = k) = h(k; N, M, n) = \begin{cases} \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{x}}; k = 0,1,2. n, \min(n, N) . \\ 0, \quad Otherwise \end{cases} \quad \text{……..(7.5.3)}$$

Where N is a positive integer, M is a positive integer not exceeding N and n is a positive integer that is at most N.

Mean and Variance of Hypergeometric Distribution:

The four moments about origin of binomial distribution are obtained as follows:

$$E(X) = \frac{nM}{N}, Var(X) = \frac{NM(N-M)(N-n)}{N^2(N-1)}$$

Recurrence Relation for the moments of Hypergeometric Distribution

$$\frac{h(k+1; N, M, n)}{h(k; N, M, n)} = \frac{(n-k)(M-k)}{(k+1)(N-M-n+k+1)}.$$

7.10.SOLVED EXAMPLE:-

Example 7.10.1.

i Comment on the following statement:
For a Binomial distribution, mean is 6 and variance is 9.

ii A die is tossed thrice. A success is getting 1 or 6 on a toss. Find the mean and variance of the number of success.

Solution:

i Mean = $\mu = np = 6$,**……………….(1)**
Variance = $npq = \sigma^2 = 9$**…………….(2)**
Dividing (2) by (1), we get
$q = \frac{9}{6}$=1.5

which is impossible as $0 \le q \le 1$.
Therefore above statement is False.

ii Probability of getting success $(1\ or\ 6)$ on a toss $=\frac{2}{6} = \frac{1}{3} = p$.
Therefore, $q = 1 - \frac{1}{3} = \frac{2}{3}$.
Number of tossed of a die, $n = 3$

a) Mean $= np = 3 \times \frac{1}{3} = 1$.
b) Variance $= npq = 3 \times \frac{1}{3} \times \frac{2}{3} = \frac{2}{3}$.

Example 7.10.2. If 10% of the bolts produced by a machine are deftive, determine the probability that out of 10 bolts chosen at random.

Solution. Here, $p(defective) = \frac{10}{100} = \frac{1}{10}$(Given).
Therefore $q(non-defective) = 1 - \frac{1}{10} = \frac{9}{10}$.
Also, $n = 10, (n\ is\ number\ of\ bolts\ chosen)$. (Given)

The probability of r defective bolts out of n bolts chosen at random is given by:

$$P(r) = \binom{n}{r} p^r q^{n-r} \ldots\ldots\ldots\ldots\ldots\ldots\ldots.. (\mathbf{1})$$

(i) Here $r = 1$,

Therefore $P(1) = \binom{10}{1}\left(\frac{1}{10}\right)^1\left(\frac{9}{10}\right)^{10-1}$

$$= \binom{10}{1}\left(\frac{1}{10}\right)^1\left(\frac{9}{10}\right)^9 = (.9)^9 = 0.3874 \ldots\ldots.. (\mathbf{2})$$

(ii) Here $r = 0$,
Therefore

$$P(0) = \binom{10}{0}\left(\frac{1}{10}\right)^0\left(\frac{9}{10}\right)^{10-0} = \left(\frac{9}{10}\right)^{10-0} = 0.3486\ldots\ldots\ldots.(3)$$

(iii) Probability that at most 2 bolts will be defective
$= P(0) + P(1) + P(2)$**…..(4)**

Now, $P(2) = \binom{10}{2}\left(\frac{1}{10}\right)^2\left(\frac{9}{10}\right)^{10-2} = 45\left[\frac{1}{10}\right]^2\left(\frac{9}{10}\right)^{10-2}$

$$= 45\left(\frac{1}{100}\right)(0.43046) = 0.1937$$

From ***(4),*** Required Probability
$= P(0) + P(1) + P(2) = 0.3486 + 0.3874 + 0.1937 = 0.9297$.

Example 7.10.3. A binomial variable X satisfies the relation $9P(X = 4) = P(X = 2)$ when $n = 6$. Find the value of the parameter p and $P(X = 1)$.

Solution: We know that,

$$P(X = r) = \binom{n}{r} p^r q^{n-r} \dots\dots\dots..(\mathbf{1})$$

Therefore $P(X = 4) = \binom{6}{4}p^4q^2 = 15p^4q^2$
and $P(X = 2) = \binom{6}{2}p^2q^4 = 15p^2q^4$. Since $n = 6$.
The given relation is

$$9P(X = 4) = P(X = 2) \Rightarrow 9(15p^4q^2) = 15p^2q^4$$

This implies that $15p^2 = q^2 = (1-p)^2$(Since$p + q = 1$)
This implies that $9p^2$= $1 + p^2 - 2p$. Therefore $8p^2 + 2p - 1 = 0$. Therefore $(4p - 1)(2p + 1) = 0$
Therefore $p = \frac{1}{4}$.
Now, $P(X = 1) = \binom{6}{1}\left(\frac{1}{4}\right)^1\left(\frac{3}{4}\right)^5 = .3559$. Since $q = \frac{3}{4}$.

Example7.10.4. Fit a binomial distribution to the following frequency data:

x	0	1	2	3	4
f	30	62	46	10	2

Solution: The table is as follows:

x	f	fx
0	30	0
1	62	62
2	46	92
3	10	30
4	02	8
	$\sum f = 15$	$\sum fx = 15$

Mean of observations = $\frac{\sum fx}{\sum f} = \frac{192}{150} = 1.28$
$\Longrightarrow np = 1.28$
$\Longrightarrow 4p = 1.28$ ($n$is no. of trial)
$\Longrightarrow p = 0.32$
Therefore $q = 1 - p = 1 - 0.32 = 0.68$
Also, $p = 0.32, q = 1 - p = 1 - 0.32 = 0.68$
Also, $N = 150$.
Hence the binomial distribution is =
$N(q + p)^n$ =150(0.68 + 0.32)^4.

Example7.10.5.A learner is given a true-false examination with8 questions. If he corrects at least 7 questions, he passes the examination. Find the probability that he will pass given that he guesses all questions.

Solution. Here,n = number of questions asked = 8.
$p = \frac{1}{2}, q = \frac{1}{2}$.(Since the question can either be true or false).

Probability that he will pass = $P(r \geq 7) = P(7) + P(8)$

$$= \binom{8}{7}\left(\frac{1}{2}\right)^7\left(\frac{1}{2}\right)^{8-7} + \binom{8}{8}\left(\frac{1}{2}\right)^7\left(\frac{1}{2}\right)^{8-8}$$

$$= 8\left(\frac{1}{2}\right)^7\left(\frac{1}{2}\right) + 1.\left(\frac{1}{2}\right)^8$$

$$=\left(\frac{1}{2}\right)^8(8+1) = \frac{9}{256} = .03516$$

Example7.10.6. Six dice are thrown729 times. How many times do you expect at least three dice to show a five or six?

Solution.p =the chance of getting 5 or 6 with one die = $\frac{2}{6}=\frac{1}{3}$.

$p = 1 - \frac{1}{3} = \frac{2}{3}, n = 6, N = 729.$

Since dice are in sets of 6 and there are 729 sets.

The expected number of times at least three dice showing five or six

$$= N.P(r \geq 3) = 729[P(3) + P(4) + P(5) + P(6)]$$

$$= 729[P(3) + P(4) + P(5) + P(6)]$$

$$=729\left[\binom{6}{3}\left(\frac{2}{3}\right)^3\left(\frac{1}{3}\right)^3 + \binom{6}{4}\left(\frac{2}{3}\right)^2\left(\frac{1}{3}\right)^4 + \binom{6}{5}\left(\frac{2}{3}\right)^1\left(\frac{1}{3}\right)^5 + \binom{6}{6}\left(\frac{1}{3}\right)^6\right]$$

$$=\frac{729}{3^6}[160 + 60 + 12 + 1] = 233.$$

Example7.10.6. The probability of a man hitting a target is $\frac{1}{3}$. How many times must be fire so that the probability of his hitting the target at least once is more than 90%?

Solution.p =the chance of getting 5 or 6 with one die = $\frac{2}{6}=\frac{1}{3}$.

The probability of not hitting the target in n trials is q^n. Therefore, to find the smallest n for which the probability of hitting at least once is more than 90%,. We have $1 - q^n > 0.9$. This implies that $1 - \left(\frac{2}{3}\right)^n > 0.9 \Rightarrow \left(\frac{2}{3}\right)^n < 0.1$.The smallest n

For which the above inequality holds true is 6 hence he must fire 6 times.

Example7.10.7. In a bombing action, there is50% chance that any bomb will strike the target. Two direct hits are needed to destroy the target completely. How many bombs are required to be dropped to give a 99%chance or better of completely destroying the target?

Solution. $p = \frac{50}{100} = \frac{1}{2}$.
Since the probability must be greater than 0.99,if . $n$bombs are dropped, we have

$$\binom{n}{2}\left(\frac{1}{2}\right)^n + \binom{n}{3}\left(\frac{1}{2}\right)^n + \binom{n}{4}\left(\frac{1}{2}\right)^n + \cdots \binom{n}{n}\left(\frac{1}{2}\right)^n \geq 0.99$$
$$\left(\frac{1}{2}\right)^n \left[\binom{n}{2} + \binom{n}{3} + \binom{n}{4} + \cdots \binom{n}{n}\right] \geq 0.99$$
$$\frac{2^n - n - 1}{2^n} \geq 0.99.$$

Example7.10.8.

(i) Suppose that a book of 600pages contains 40 printing mistakes. Assume that these errors are randomly distributed throughout the book and x, the number of errors per page has a Possion distribution. What is the probability that 10 pages selected at random will be free from errors?

(ii) Wireless sets are manufactured with 25 solders joints each, on the average 1 joint in 500 is defective. How many sets can be expected to be free from defective joints in a consignment of 1000 sets?

Solution. (i) $p = \frac{40}{600} = \frac{1}{15}$, $\quad n = 10$.
$\lambda = np = 10\left(\frac{1}{15}\right) = \frac{2}{3}$.

$$P(x) = \frac{e^{-\lambda}\lambda^x}{x!} = \frac{e^{-\frac{2}{3}}\frac{2}{3}^x}{x!}$$

$$P(r) = \frac{e^{-\lambda}\lambda^x}{x!} = \frac{e^{-\frac{2}{3}}(2/3)^x}{x!}$$

Therefore, $P(r) = \frac{e^{-\lambda}\lambda^x}{x!} = \frac{e^{-\frac{2}{3}}(2/3)^0}{x!} = e^{-2/3} = 0.51$
(ii)$p = \frac{1}{500}, n = 25,$
Therefore $\lambda = np = 10\left(\frac{1}{500}\right) = \frac{1}{20} = 0.05.$
Number of sets in a consignment, $N = 10000$.
Probability of having no defective joint = , $P(r = 0) = \frac{e^{-.05}(0.05)^0}{0!} = 0.9512.$
Therefore the expected number of sets free from defective joints $= 0.9512 \times 10000 = 9512.$

Example7.10.9. A car – hire firm has two cars, which it hires out day by day. The number of demands for a car on each day is distributed as a Poisson distribution with mean 1.5. Calculate the proportion of days on which neither car is used and the proportion of days on which some demand is refused($e^{-1.5} = 0.2231$).

Solution. Since the number of demands for a car is distributed as a Poisson distribution with $\lambda = 1.5$.
Therefore, proportion of days on which neither car is used = Probability of there being no demand for the car.

$$= \frac{\lambda^0 e^{-\lambda}}{0!} = e^{-1.5} = 0.2231.$$

Proportion of days on which some demand is refused = Probability for the number of demands to be more than two

$$= 1 - P(x \leq 2) = 1 - \left(e^{-\lambda} + \frac{\lambda e^{-\lambda}}{1!} + \frac{\lambda^2 e^{-\lambda}}{2!}\right)$$

$$= 1 - e^{-1.5}\left(1 + \frac{1.5}{1!} + \frac{(1.5)^2}{2!}\right) = 0.1912625$$

Example7.10.10. An insurance company finds that0.005% of the population dies from a certain kind of accident each year. What is the probability that the company must pay off no more than 3 of 10,000 insured risks against such incident in a given year?

Solution.$p = \frac{0.005}{100} = 0.00005, n = 10000$

Therefore $\lambda = np = 10000 \times 0.00005 = 0.5$
Required Probability = $1 - P(r \leq 3) = 1 - [P(0) + P(1) + P(2) + P(3)]$

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

$$= 1 - \left[\frac{e^{-.05}(0.05)^0}{0!} + \frac{e^{-.05}(0.05)^1}{1!} + \frac{e^{-.05}(0.05)^2}{2!} + \frac{e^{-.05}(0.05)^3}{3!}\right]$$

$$= 1 - e^{-.05}[1 + 0.5 + 0.125 + 0.021] = 0.0016$$

Example7.10.11. (i) Six coins are tossed 6400 times. Using the Poisson distribution, determine the approximate probability of getting six heads x times.

(ii) A Poisson distribution has a double mode at $x = 3$ and $x = 4$.What is the probability that x will have one or the other of these two values?

Solution.
(i) Probability of getting one head with one coin = $\frac{1}{2}$.
Therefore the probability of getting six heads with six coins = $\left(\frac{1}{2}\right)^6 = \frac{1}{64}$.

Therefore average number of six heads with six coins in 6400 throws = $np = 6400 \times \frac{1}{64} = 100$. Therefore the mean of the Poisson distribution = 100.
Approximate probability of getting six heads x times when the distribution is Poisson = $\frac{\lambda^x e^{-\lambda}}{x!} = \frac{(100)^x . e^{-100}}{x!}$

(ii) Since 2 modes are given when λ is an integer, modes are $\lambda - 1$ and λ.
Therefore $\lambda - 1 = 3 \Rightarrow \lambda = 4$
Probability (when$r = 3$)=$\frac{e^{-.4}(4)^3}{3!}$
Probability (when$r = 4$)=$\frac{e^{-.4}(4)^4}{4!}$
Required probability = P($r = 3$ or 4) = P($r = 3$) + P($r = 4$)

$$= \frac{e^{-.4}(4)^3}{3!} + \frac{e^{-.4}(4)^4}{3!} = \frac{64}{3} e^{-4} = 0.39073.$$

Example7.10.12. Given the hypothetical distribution:

No. of cells (x)	0	1	2	3	4	5	Total
Frequency (f)	213	128	37	18	3	1	400

Fit a negative binomial distribution and calculate the expected frequencies.

Solution. Let X be negative binomial variate with parameters r and p.
$\mu_1' =$ Mean $= \frac{\sum fx}{\sum f} = \frac{273}{400} = 0.6825 = \frac{rq}{p}; (q = 1 - p)$**….(7.10.12.1)**
$\mu_1' = \frac{\sum fx^2}{\sum f} = \frac{511}{400} = 1.2775. \mu_2 = \mu_2' - {\mu_1'}^2 = 1.2775 - (0.6825)^2 = 0.8117$
Therefore Variance = 1.2775 $= \frac{rq}{p^2}$**………..(7.10.12.2)**
Dividing (7.10.12.1) by (7.10.12.2) we get

$$p = \frac{0.6825}{0.8117} = 0.8408, q = 1 - p = 0.1592$$

Therefore, $p = \frac{p \times 0.6825}{q} = \frac{0.5738}{0.1592} = 3.6043 \approx 4$
Since, r being the number of successes cannot be fractional.$f_0 = p^r = (.8408)^4 = 0.4978 \approx 0.5$

$$f_1 = \frac{r+0}{0+1} q f_0 = rqf_0 = 0.5738 \times 0.5 = 0.2869$$

$$\therefore rq = p \times 0.6825 = 0.5738$$

$$f_2 = \frac{r+1}{1+1}.q.f_1 = \frac{5}{2} \times 0.1592 \times 0.2869 = 0.1142$$
$$f_3 = \frac{r+2}{2+1}.q.f_2 = \frac{6}{3} \times 0.1592 \times 0.2869 = 0.0364$$
$$f_4 = \frac{r+3}{3+1}.q.f_3 = \frac{7}{4} \times 0.1592 \times 0.0364 = 0.0101$$
$$f_5 = \frac{r+4}{4+1}.q.f_4 = \frac{8}{5} \times 0.1592 \times 0.0101 = 0.0026$$

Therefore expected frequencies are :($N = 400$).

Nf_0	Nf_1	Nf_2	Nf_3	Nf_4	Nf_5
200	114.76	45.68	14.56	4.04	1.04

Observed frequency	213	128	37	18	3	1
Expected Frequency	200	115	46	14	4	1

Example7.10.13. Suppose X is a non-negative integral valued random variable.

Show that the distribution of X is geometric if it 'lacks memory', if for each $k \geq 0$ and $Y = X - k$, one has $P(Y = t/X \geq k) = P(X = t), for\ t \geq 0$.

Solution. Let us suppose$P(X = r) = p_r; r = 0,1,2, ...$

Define

$q_k = P(X \geq k) =$
$p_k + p_{k+1} + p_{k+2} \dots \dots \dots \dots \dots \dots \dots . . (\mathbf{7.10.13.1})$

We are given:

$P(Y = t/X \geq k) = P(X = t) = p_t \dots \dots . . (\mathbf{7.10.13.2})$

We have

$P(Y = t/X \geq k) = \frac{P(Y=t/X\geq k)}{P(X\geq k)} = \frac{P(X-k=t \cap X\geq k)}{P(X\geq k)} = \frac{P(X=k+t)}{P(X\geq k)} =$
$\frac{p_{k+t}}{q_k}$ for every$t \geq 0$ and all $k \geq 0$[From($\mathbf{7.10.13.2}$)].

In particular, taking $k = 1$, we get

$p_{t+1} = q_1.p_t = (p_1 + p_2 + \cdots)p_t = (1 - p_0)p_t$[From($\mathbf{7.10.13.1}$)]

This implies that $p_t = (1 - p_0)p_{t-1} = (1 - p_0)^2 p_{t-2} = \cdots = (1 - p_0)^t p_0$.

Hence $p_t = P(X = t) = p_0(1 - p_0)^t; t = 0,1,2 \dots$

This implies X has a geometric distribution.

Example7.10.14. Explain how you will use hypergeometric model to estimate the number of fish in a lake.

Proof. Let us suppose that in a lake there are N fish, N unknown. The problem is to estimate N. A catch of $'r'$ fish (all at the same time) is made and these fish are returned alive into the lake after making each with a red spot. After a reasonable period of time, a during which these 'marked' fish are assumed to have distributed themselves 'at random' in the lake, another catch of 's' fish (again, all at once) is made. Here r and s are regarded as fixed pre-determined constants. Among these s fish caught, there will be (say) X is a random variable following discrete probability function given by hypergeometric model:

$$f_X(x/N) = \frac{\binom{r}{x}\binom{N-r}{S-x}}{\binom{N}{x}} = p(N) \dots\dots\dots.(\mathbf{7.10.14.1})$$

Where x is an integer such that $max(0, s - N + r) \leq$ min (r, s) and $f_X(x/N) = 0$ Otherwise.The value of N is estimated by the principle of Maximum Liklihood (In the unit of theory of estimation)*i.e.,*the principle of maxima and minima in calculus cannot be used here. Here we proceed as follows:$\lambda(N) = \frac{p(N)}{p(N-1)} = \frac{(N-r)(N-s)}{N(N-r-s+x)}$

Therefore $\lambda(N) > 1$ iff $N > \frac{rs}{x} \Rightarrow p(N) > p(N-1)$ iff $N > \frac{rs}{x}$..$(\mathbf{7.10.14.2})$

And $\lambda(N) > 1$iff$N < \frac{rs}{x} \Rightarrow p(N) > p(N-1)$ iff $N < \frac{rs}{x}$ ....$(\mathbf{7.10.14.3})$

From $(\mathbf{7.10.14.2})$ and $(\mathbf{7.10.14.3})$ we see that $f_X(x/N) = p(N)$ reaches the maximum value when N is approximately equal to $\frac{rs}{x}$. Hence maximum likelihood estimate of function N is given by $\widehat{N}(X) = \frac{rs}{x}$ it implies that $\widehat{N}(X) = \frac{rs}{X}$.

CHECK YOUR PROGRESS

Problem1. In a Binomial Distribution, if 'n' is the number of trials and 'p' is the probability of success, then the mean value is given by ___________

a) np
b) n
c) p
d) $np(1-p)$

Problem2. In a Binomial Distribution, if p, q and n are probability of success, failure and number of trials respectively then variance is given by ___________

a) npq
b) npq
c) nq
d) *none of the above*

CHECK YOUR PROGRESS

Problem 3. In a Poisson Distribution, if 'n' is the number of trials and 'p' is the probability of success, then the mean value is given by?
a) $m = np$
b) $m = np^2$
c) $m = np(1-p)$
d) $m = p$

Problem 4. If 'm' is the mean of a Poisson Distribution, then variance is given by ___
a)$m/3$
b)$m^{1/2}$
c) m
d) $m/2$
Problem 5. To construct a binomial probability distribution, the mean must be known. True/False

Problem 6. The Poisson probability distribution is a continuous probability distribution. True/False

Problem 7. In Bernouli distribution trials are independent of each other True\False

Problem 8. The geometric distribution is superb for understanding when an event might first occur. True\False

Problem 9. Inhypergeometric distribution the result of each draw (the elements of the population being sampled) can be classified into one of two mutually exclusive categories. True\False.

Problem 10. If we define an example rolling a 6 on a dice as a success, and rolling any other number as a failure, and ask how many failure rolls will occur before we see the third success $(r = 3)$In such a case, the probability distribution of the number of failures that appear will be a negative binomial distribution. True\False

7.11SUMMARY:-

In this unit we have covered the topic Discrete uniform Distribution, .Bernoulli Distribution, Binomial Distribution, Possion Distribution. Negative Binomial Distribution. Geometric Distribution and Hyper geometric Distribution.

7.12 GLOSSARY:-

i Random variable
ii Variance
iii Moments
iv Mathematical expectation
v Moment generating function.

7.13. REFERENCE BOOKS:-

1. S. C. Gupta and V. K. Kapoor, (2020), *Fundamentals of mathematical statistics*, Sultan Chand & Sons.
2. Seymour Lipschutz and John J. Schiller, (2017), *Schaum's Outline: Introduction to Probability and Statistics*, McGraw Hill Professional.
3. J. S. Milton and J. C. Arnold , (2003*), Introduction to Probability and Statistics (4th Edition),* Tata McGraw-Hill.
4. https://www.wikipedia.org.

7.14.SUGGESTED READINGS:-

1. A.M. Goon,(1998), *Fundamental of Statistics (7th Edition),* 1998.
2. R.V. Hogg and A.T. Craig, (2002), *Introduction to Mathematical Statistics,* MacMacMillan, 2002.
3. Jim Pitman, (1993), *Probability,* Springer-Verlag.
4. https://archive.nptel.ac.in/courses/111/105/111105090.

7.15.TERMINAL QUESTIONS:-

i. During war, 1 ship out of 9 was sunk on an average in making a certain voyage. What was the probability that exactly 3 out of a convoy of 6 ships would arrive safely?

ii. A policeman frees 6 bullets on a dacoit. The probability that the dacoit will be killed by a bullet is 0.6. What is the probability that dacoit is still alive?

iii. If the probability of hitting a target is 10% and 10 shots are fired independently. What is the probability that the target will be hit at least once?

iv. Out of 800 families with 4 children each, how many families would be expected to have (i) 2 boys and 2 girls (ii) at least one boy (iii) no girl (iv) atmost two girl? Assume equal probabilities for boys and girls?

v. Show that in a Poisson distribution with unit mean, mean deviation about mean $\left(\frac{2}{e}\right)$ is times the standard deviation.

vi. In a certain factory manufacturing razor blades, there is a small chance of 0.002 for any blade to be defective. The blades are supplied in packets of 10. Use suitable distribution to calculate the approximate number of packets containing no defective, one defective and two defective respectively in a consignment of 20,000 packets.

vii. Let X_1, X_2 be independent random variables each having geometric Distribution $q^k p: k = 0,1,2, \ldots \ldots$...Show that the conditional distribution of X_1 given $X_1 + X_2$ is uniform.

viii. Obtain the Poisson distribution as a limiting case of the negative binomial distribution?

ix. Describe the probability model from which the binomial distribution can be generated. Hence find the first four central moments……………………

x. Obtain the Moment generating function of the Binomial Distribution…………

7.16.ANSWERS:-

Answer of Check your progress Questions:-

CHQ1:(a)np.
CHQ2:(b)npq.
CHQ3: (a) np
CHQ4: (c) m
CHQ5: True
CHQ6: False
CHQ7: True
CHQ8: True
CHQ9: True
CHQ10: True

Answer of Terminal Questions:-

TQ I: $\frac{10240}{9^6}$.
TQ II: $.004096$.
TQ III: 0.6513.
TQ IV: 550.
TQVI: (a)53.84%(b)19604,392,4 & 0 $packets$.

UNIT :-8 CONTINUOUS PROBABILITY DISTRIBUTIONS

CONTENTS:

8.1. INTRODUCTION:-

In previous unit we have discussed about Discrete uniform Distribution. In the discrete distribution we have explained about Bernoulli Distribution, Binomial Distribution, Possion Distribution, Negative Binomial Distribution, Geometric Distribution and Hypergeometric Distribution. Now in present unit we are explaining about continuous probability distribution. In the continuous probability distribution our main focuses on Normal distribution. After normal distribution we are defining and explaining the central limit theorem.

The beginning of the normal distribution is very interesting [Stigler, 1986]. In the research work of Abraham DeMoivre in the mid-18^{th} century the concept of normal distribution was started and then Gauss extended the work in the late 18^{th} and early 19^{th} centuries Gauss, who first made reference to it in the beginning of 19^{th} century ***(1809),*** as the distribution of errors in Astronomy.

Gauss used the normal curve to describe the theory of accidental errors of measurement involved in the calculation of orbits of heavenly bodies.

Johann Carl Friedrich Gauss

(30-04- 1777 – 23-02- 1855)

Fig 8.1.1

Ref: https://en.wikipedia.org/wiki/Carl_Friedrich_Gauss

.

8.2.OBJECTIVES:-

After studying this unit learner will be able to:

1. Describe the notion of Continuous probability distribution.
2. Explain the Normal distribution.
3. Understand the Central limit theorem.

8.3.NORMAL DISTRIBUTION:-

A random variable X is said to have a normal distribution with parameters $\mu(called\ 'mean')$ and $\sigma^2(called'variance')$ if its probability density function is given by the probability law:

$$f(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}$$

or

$$f(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/2\sigma^2}; -\infty < x < \infty, -\infty < \mu < \infty,$$

$$\sigma > 0 \ldots\ldots\ldots\ldots\ldots\ldots\ldots. (\mathbf{8.3.1})$$

- In simple words X is distributed as $N(\mu, \sigma^2)$ and is expressed by $X \sim N(\mu, \sigma^2)$.
- If $X \sim N(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma}$, is a standard normal variate with $E(Z) = 0$ and $\text{Var}(Z) = 0$ and we write $Z \sim N(0,1)$.
- The probability density function $(p.d.f)$ of standard variate$Z$ is given by:
$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, -\infty < z < \infty,$$
and the corresponding distribution function, denoted by
$\Phi(z) = P(Z \le z) = \int_{-\infty}^{z} \varphi(u)\, du = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \varphi(u) du = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{u^2/2}\, du.$
Now the two important results on the distribution function $\Phi(.)$ of standard normal variate.

- $\Phi(-z) = 1 - \Phi(z), z > 0.$

Proof.$\Phi(-z) = P(Z \le -z) = P(Z \ge z) = 1 - P(Z \le z) = 1 - \Phi(z).$

- $P(a \le X \le b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$, where $X \sim N(\mu, \sigma^2)$.

Proof. $P(a \le X \le b) = P\left(\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma}\right), \quad \left(Z = \frac{X-\mu}{\sigma}\right)$

$$= P\left(Z \le \frac{b-\mu}{\sigma}\right) - P\left(Z \le \frac{a-\mu}{\sigma}\right) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right).$$

8.3.1.NORMAL DISTRIBUTION AS A LIMITING FORM OF BINOMIAL DISTRIBUTION:-

Normal distribution is another limiting form of the binomial distribution under the following conditions:

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

ii neither$p$ nor q is very small.

The $p.m.f.$ of the binomial distribution with parameters n and p is given by:

$$p(x) = \binom{n}{x} p^x q^{n-x} = \frac{n!}{x!\,(n-x)!} p^x q^{n-x}; x = 0,1,2, \ldots \ldots n \ldots \ldots \ldots \ldots . (\mathbf{8.3.1})$$

Let us now consider the standard binomial variate:

$$Z = \frac{X - E(X)}{\sqrt{V(X)}} = \frac{X - np}{\sqrt{npq}}; X = 0,1,2, \dots . . n \dots \dots \dots \dots \dots \dots . . (\mathbf{8.3.2})$$

When $X = 0, Z = \frac{-n}{\sqrt{npq}} = -\sqrt{\frac{np}{q}}$ and when $X = n, Z = \frac{n-np}{\sqrt{npq}} = \sqrt{\frac{nq}{p}}$

Thus in the limit as $n \to \infty, Z$ takes the values $-\infty$ to ∞. Hence the distribution of X will be a continuous distribution over the range $-\infty$ to ∞.

We want the limiting form of $(\mathbf{8.3.1})$ under the above two conditions. Using Stirling's approximation to $r!$ For larger , $viz.$, $\lim_{r\to\infty} r! \simeq \sqrt{2\pi}\, e^{-r} r^{r+(1/2)}$,

$$\lim p(x) = lim \left[\frac{\sqrt{2\pi}\, e^{-n} n^{n+\frac{1}{2}} p^x q^{n-x}}{\sqrt{2\pi}\, e^{-x} x^{x+\frac{1}{2}} \sqrt{2\pi}\, e^{n-x} (n-x)^{n-x+\frac{1}{2}}}\right]$$

$$= lim \left[\frac{1}{\sqrt{2\pi}} . \frac{1}{\sqrt{npq}} \frac{(np)^{x+\frac{1}{2}} (nq)^{n-x+\frac{1}{2}}}{x^{x+\frac{1}{2}} (n-x)^{n-x+\frac{1}{2}}}\right]$$

$$= lim \left[\frac{1}{\sqrt{2\pi}} . \frac{1}{\sqrt{npq}} \left(\frac{np}{x}\right)^{x+\frac{1}{2}} \left(\frac{np}{n-x}\right)^{n-x+\frac{1}{2}}\right] \dots \dots \dots . (\mathbf{8.3.3})$$

From ***(8.3.2),*** we get $X = np + Z\sqrt{npq} \Rightarrow \frac{X}{np} = 1 + Z\sqrt{\frac{q}{np}}$

Further

$n - X = n - np - Z\sqrt{npq} = nq - Z\sqrt{npq} \Rightarrow \frac{n-X}{nq} = 1 - Z\sqrt{\frac{p}{nq}}$.

Also $dz = \frac{1}{\sqrt{npq}} dx$

Hence the probability differential of the distribution of Z, in the limit is:

$$dG(z) = g(z)dz = \lim_{n\to\infty} \left(\frac{1}{\sqrt{2\pi}} \times \frac{1}{N}\right) dz,$$

where $N = \left(\frac{x}{np}\right)^{x+\frac{1}{2}} \left(\frac{n-x}{nq}\right)^{n-x+\frac{1}{2}}$ ……………………………………(8.3.4)

This implies $logN = \left(x + \frac{1}{2}\right) log(x/np) + \left(n - x + \frac{1}{2}\right) log\{(n-x)/nq\}$

$$= \left(np + z\sqrt{npq} + \frac{1}{2}\right) log\left\{1 + z\sqrt{q/np}\right\} + \left(nq - z\sqrt{npq} + \frac{1}{2}\right) log\left\{1 - z\sqrt{(p/nq)}\right\}$$

$$= \left(np + z\sqrt{npq} + \frac{1}{2}\right)\left\{-z\sqrt{(p/nq)} - \frac{1}{2}z^2(p/nq) - \frac{1}{3}z^3(p/nq)^{3/2} - \cdots\right\}$$

$$= \left[\left\{z\sqrt{npq} - \frac{1}{2}qz^2 + \frac{1}{3}z^3\frac{q^{3/2}}{\sqrt{np}} + z^2q + \frac{1}{2}z^3q - \frac{1}{2}z^3\frac{q^{3/2}}{\sqrt{np}} + \frac{1}{2}z\sqrt{\frac{q}{np}} - \frac{1}{4}z^2\frac{p}{np} + \cdots\right\}\right]$$

$$= \left[-\frac{1}{2}z^2(p+q) + z^2(p+q) + \frac{z}{2\sqrt{n}}\left(\frac{q}{p} + \frac{p}{q}\right) + 0\left(n^{-1/2}\right)\right]$$

$= \frac{1}{2}z^2 + 0\left(n^{-1/2}\right) \longrightarrow \frac{z^2}{2} \text{as} n \to \infty,$

Therefore $\lim_{n\to\infty} \log N = \frac{z^2}{2} \Longrightarrow \lim_{n\to\infty} \log N = e^{z^2/2}$.

Substituting in (8.3.4), we get

$dG(z) = g(z)dz = \frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}, -\infty < z < \infty$ ……………… ***(8.3.5)***

Hence the probability function of Z is:

$$g(z) = \frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}, -\infty < z < \infty \text{ ……………………. } \mathbf{(8.3.6)}$$

This is the probability density fuction of the normal distribution with mean 0
and unit variance.

If X is a normal variate with mean μ and standard deviation σ, then $Z = (X - \mu)/\sigma$, is standard normal variate. Jacobian of transformation is $1/\sigma$. Hence substituting in ***(8.3.6)***, the p.d.f. of a normal variate$X$ with $E(X) = \mu, Var(X) = \sigma^2$ is given by :

$$f_X(x) = \begin{cases} \frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/2\sigma^2}, -\infty < x < \infty \\ 0 \;\; otherwise \end{cases}$$

- Normal distribution can also be obtained as a limiting case of Poisson distribution with the parameter $\lambda \to \infty$.

8.3.2.NORMAL DISTRIBUTION AND NORMAL PROBABILITY CURVE:-

The normal probability curve with mean μ and standard deviation σ is given by the equation:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}, -\infty < x < \infty$$

and has the following properties:

(i) The curve is bell-shaped and symmetrical about the line $X = \mu$.

(ii) Mean, median and mode of the distribution coincide.

(iii) As x increases numerically, $f(x)$ decreases rapidly, the maximum probability occurring at the point $x = \mu$, and is given by : $[p(x)]_{max} = \frac{1}{\sigma\sqrt{2\pi}}$.

(iv) $\beta_1 = 0, \beta_2 = 3$.

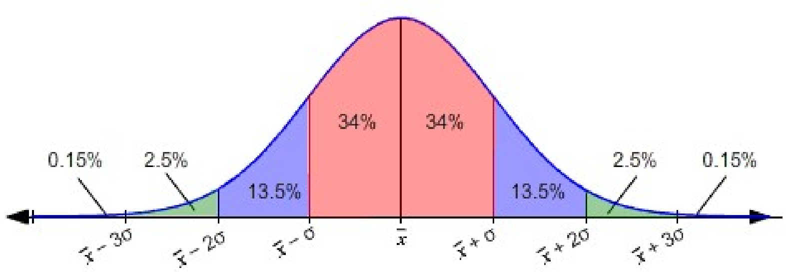
(v) $\mu_{2r+1} = 0, (r = 0,1,2 \ldots\ldots)$
And $\mu_{2r} = 1,3,5 \ldots\ldots (2r-1)\sigma^{2r}, (r = 0,1,2, \ldots\ldots)$.

(vi) Since$f(x)$ being the probability, can never be negative, no portion of the curve lies below the x −axis.

(vii) Linear combination of independent normal variates is also a normal variate.

(viii) x −axis is an asymptote to the curve.

(ix) The points of inflexion of the curve are :$x = \mu \pm \sigma, f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2}$



$X = \mu$

Normal Probability Curve

Fig 8.3.2

Ref:
https://www.varsitytutors.com/hotmath/hotmath_help/topics/normal-distribution-of-data

(x) Mean deviation about mean = $\sqrt{\frac{2}{\pi}}\sigma \simeq \frac{4}{5}\sigma$ (approx.)

(xi) Quartiles are given by: $Q_1 = \mu - 0.6745\sigma$; $Q_3 = \mu + 0.6745\sigma$

(xii) $Q.D = \frac{Q_3 - Q_1}{2} \simeq \frac{2}{3}\sigma$. We have (approximately).

$$Q.D: M.D.: S.D. :: \frac{2}{3}\sigma : \frac{4}{5}\sigma : \sigma :: \frac{2}{3} : \frac{4}{5} : 1$$

$$\Rightarrow Q.D: M.D.: S.D. 10: 12: 15$$

(xiii) Area Property :

$P(\mu - \sigma < X < \mu + \sigma) = 0.6826,\quad P(\mu - 2\sigma < X < \mu + 2\sigma) = 0.9544,$

and $P(\mu - 3\sigma < X < \mu + 3\sigma) = 0.9973$.

The adjoining 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 $\pm\sigma$	***Area under the curve***
$Z = \pm 0.745$ $Z = \pm 1.000$ $Z = \pm 1.96$ $Z = \pm 2.00$ $Z = \pm 2.58$ $Z = \pm 3.00$	$50\% = 0.50$ $68.26\% = 0.6826$ $95\% = 0.95$ $95.44\% = 0.9544$ $99\% = 0.99$ $99.73\% = 0.9973$

(xiv) If X and Y are independent standard normal variates, then it can be easily proved that $U = X + Y$ and $V = X - Y$ are independently distributed, $U = X + Y$ and $V = X - Y$ are independently distributed, $U \sim N(0,2)$ and $V \sim N(0,2)$.

Bernstein's Theorem.If X and Y are independent and identically distributed random variables with finite variances and if $U = X + Y$ and $V = X - Y$ are independent, then all random variable X, Y, U and V are normally distributed.

(xv) We state below another result which characterises the normal distribution.

If $X_1, X_2 \dots \dots \dots \dots X_n$ are i.id. random variables with finite variance, then the common distribution is normal if and only if :
$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $s^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ or $\sum_{i=1}^{n} X_i$ and $\sum_{i=1}^{n}(X_i - \bar{X})^2$ are independent.

8.3.3.PROPERTIES OF NORMAL DISTRIBUTION:-

Mode of Normal Distribution.

- $x = \mu$ is the mode of the normal distribution.

Median of Normal Distribution.

- For the normal distribution, Mean = Median.
- For the normal distribution mean, median and mode coincide. Hence the distribution is symmetrical.

Moment generating of Normal Distribution.

The $m.g.f.(about\ origin)$ is given by:

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f(x)dx$$

$$= \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{tx} exp\{-(x-\mu)^2/2\sigma^2\}\, dx$$

$$= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} exp\{t(\mu+\sigma z)\}exp(z^2/2)\, dz, \left(z = \frac{x-\mu}{\sigma}\right)$$

$$= e^{\mu t}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} exp\left\{-\frac{1}{2}(z^2 - 2t\sigma z)\right\} dz$$

$$= e^{\mu t}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} exp\left[-\frac{1}{2}\{(z-\sigma t)^2 - \sigma^2 t^2\}\right]$$

$$= e^{\mu t +} t^2\sigma^2/2 \times \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}(z-\sigma t)^2\right] dz$$

$$= e^{\mu t +} t^2\sigma^2/2 \times \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} exp(-u^2/2)\, du$$

Hence $M_X(t) = e^{\mu t + t^2\sigma^2/2}$**…………………*(8.3.3)***

- $M.G.F.$ of standard Normal Variate. If $X \sim N(\mu, \sigma^2)$, then standard normal variate is given by: $Z = (X-\mu)/\sigma$.
  $M_z(t) = e^{-\mu t/\sigma} M_X(t/\sigma)$
  $= exp(-\mu t/\sigma).exp\{(\mu t/\sigma)$
  $+ (t^2/\sigma^2)(\sigma^2/2)\}$
  $= exp(t^2/2) \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots (\mathbf{8.3.4})$
- Similarly $Z \sim N(0,1)$. Hence $\mu = 0$ and $\sigma^2 = 1$ in (8.3.3), we get
  $M_z(t) = exp(t^2/2)$.

Moments of Normal Distribution.

Odd order moments about mean are given by:

$$\mu_{2n+1} = \int_{-\infty}^{\infty} (x-\mu)^{2n+1} f(x)dx$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x - \mu)^{2n+1} \exp\{-(x-\mu)^2/2\sigma^2\} dx$$

Therefore

$$\mu_{2n+1} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma z)^{2n+1} \exp\{-z^2/2\}\, dz, \text{ Where } z = \frac{x-\mu}{\sigma}$$

$$= \frac{\sigma^{2n+1}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^{2n+1} \exp\{-z^2/2\}\, dz = 0, .. (8.3.5).$$

Since the integrand $z^{2n+1}e^{-z^2/2}$ is an odd function of z.

Even order moments about mean are given by:

$$\mu_{2n} = \int_{-\infty}^{\infty} (x-\mu)^{2n} f(x)dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma z)^{2n} \exp\{-z^2/2\} dx$$

$$= \frac{\sigma^{2n}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^{2n} \exp\{-z^2/2\}\, d$$

$$= \frac{\sigma^{2n}}{\sqrt{2\pi}} 2 \int_{0}^{\infty} z^{2n} \exp\{-z^2/2\}\, dz$$

(Since integrand is an even function of z.)

$$= \frac{2\sigma^{2n}}{\sqrt{2\pi}} \int_{0}^{\infty} (2t)^n e^{-t} \frac{dt}{\sqrt{2t}}, \left(t = \frac{z^2}{2}\right)$$

Therefore $\mu_{2n} = \frac{2^n\sigma^{2n}}{\sqrt{\pi}} \int_0^\infty e^{-t}\, t^{\left(n+\frac{1}{2}\right)-1} dt$

$\Longrightarrow \mu_{2n} = \frac{2^n\sigma^{2n}}{\sqrt{\pi}} \Gamma\left(n + \frac{1}{2}\right).$

Changing n to $(n-1)$, we get

$$\mu_{2n-2} = \frac{2^{n-1}\sigma^{2n-2}}{\sqrt{\pi}} \Gamma\left(n - \frac{1}{2}\right)$$

$$\therefore \frac{\mu_{2n}}{\mu_{2n-2}} = 2\sigma^2 . \frac{\Gamma\left(n+\frac{1}{2}\right)}{\left(n-\frac{1}{2}\right)} = 2\sigma^2\left(n - \frac{1}{2}\right) [\because (r-1)\Gamma(r-1)]$$

This implies that

$$\mu_{2n} = \sigma^2(2n-1)\mu_{2n-2} \ldots\ldots\ldots. (\mathbf{8.3.6})$$

Which gives the recurrence relation for the moments of normal distribution. From (8.3.6), we have

$$\mu_{2n} = [(2n-1)\sigma^2][(2n-3)\sigma^2][(2n-5)\sigma^2]\mu_{2n-6}$$

$$= [(2n-1)\sigma^2][(2n-3)\sigma^2][(2n-5)\sigma^2] \ldots\ldots\ldots (3\sigma^2)(1\sigma^2).\mu_0$$

$$= 1.3.5 \ldots\ldots. (2n-1)\sigma^{2n} \ldots\ldots\ldots (\mathbf{8.3.7})$$

From **(8.3.6)** and **(8.3.7)** implies that for the normal distribution all odd order moments about mean vanish and even order moments about mean are given by **(8.3.7).**

A linear combination of independent normal variates is also a normal variate.
Let $X_i, (i = 1,2,3, \dots \dots, n)$ be n independent normal variates with mean μ and variance σ_i^2respectively. Then $M_{X_i}^{(t)} = exp\{\mu_i t + (t^2\, \sigma_i^2/2)\}$**……………………………..(8.3.8)**
The $m.g.f.$ of their linear combination $\sum_{i=1}^{n} a_i\, X_i$,where $a_1, a_2,\dots\dots\dots a_n$ are constants, is given by:

$$M_{\sum_i a_i X_i}(t) = \prod_{i=1}^{n} M_{a_i X_i}(t)(\because X_i's\ are\ independent)$$

$= M_{X_1}(a_1 t).M_{X_2}(a_2 t) \dots \dots \dots . . M_{X_n}(a_n t)$**………(8.3.9)**

$$[\because M_{cX}(t) = M_X(ct)]$$

From (8.3.9), we have $M_{X_i}(a_i t) = \ e^{\mu_i a_i t + t^2 a_i^2 \sigma_i^2/2}$
Therefore, $M_{\sum_i a_i X_i}(t) = \left[e^{\mu_1 a_1 t + t^2 a_1^2 \sigma_1^2/2} \times e^{\mu_2 a_2 t + t^2 a_2^2 \sigma_2^2/2} \times \dots \dots . \times e^{\mu_n a_n t + t^2 a_n^2 \sigma_n^2/2}\right]$ (From (8.3.9))

$$= exp\left[\left(\sum_{i=1}^{n} a_i \mu_i\right) t + t^2 \left(\sum_{i=1}^{n} a_i^2 \sigma_i^2\right)/2\right],$$

Which is the $m.g.f.$ of a normal variate with mean $\sum_{i=1}^{n} a_i \mu_i$ and variance $\sum_{i=1}^{n} a_i^2 \sigma_i^2$.
Hence by uniqueness theorem of $m.g.f.$,

$(\sum_{i=1}^{n} a_i\, X_i) \sim N[\sum_{i=1}^{n} a_i\, \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2]$**………..(8.3.10)**

- Sum as well as the difference of two independent normal variates is also a normal variate.
- Sum of independent normal variates is also a normal variate.
- If $X_i, (i = 1,2,3, \dots \dots, n)$ are identically and independently distributed as $N(\mu, \sigma^2)$and we take a_1=a_2=………..= a_n=$1/n$, then $\bar{X} \sim N(\mu, \sigma^2/n)$, where$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

- If $X_i, (i = 1,2,3, \dots \dots, n)$, are identically and independently distributed Normal variates with μ and variance σ^2, then their mean $\bar{X}$ is also $N(\mu, \sigma^2/n)$.

- The points of inflexion of the normal curve are given by $x = \mu \pm \sigma, f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2}$.
- Mean Deviation about mean $= \frac{2\sigma}{\sqrt{2\pi}} \int_0^\infty |z|\, e^{-z^2/2} dz$.
  Since in $[0, \infty], |z| = z$,
  Mean Deviation about mean $= \frac{4}{5}\sigma(approax)$.

8.3.4 .AREA PROPERTY (NORMAL PROBABILITY INTEGRAL):-

If $X \sim N(\mu, \sigma^2)$, then the probability that random value of X will lie between $X = \mu$ and $X = x_1$ is given by:

$P(\mu < X < x_1)$

$$= \int_\mu^{x_1} f(x)dx = \frac{1}{\sigma\sqrt{2\pi}} \int_\mu^{x_1} e^{-(x-\mu)^2/2\sigma^2}\, dx$$

Put $\frac{X-\mu}{\sigma} = Z \Rightarrow X - \mu = \sigma Z$

When $X = \mu, Z = 0$ and when $X = x_1, Z = \frac{x_1-\mu}{\sigma} = z_1$,

Therefore $P(\mu < X < x_1) = P(0 < Z < z_1) = \frac{1}{\sqrt{2\pi}} \int_0^{z_1} e^{-z^2/2} dz = \int_0^{z_1} \varphi(z) dz$

where $\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$, is the probability function of standard normal variate.The definite integral $\int_0^{z_1} \varphi(z) dz$ is known as normal probability integral and gives the area under standard normal curve between the ordinates at $Z = 0$ and $Z = z_1$.These area have been tabulated for different values of z_1, at interval of 0.01.

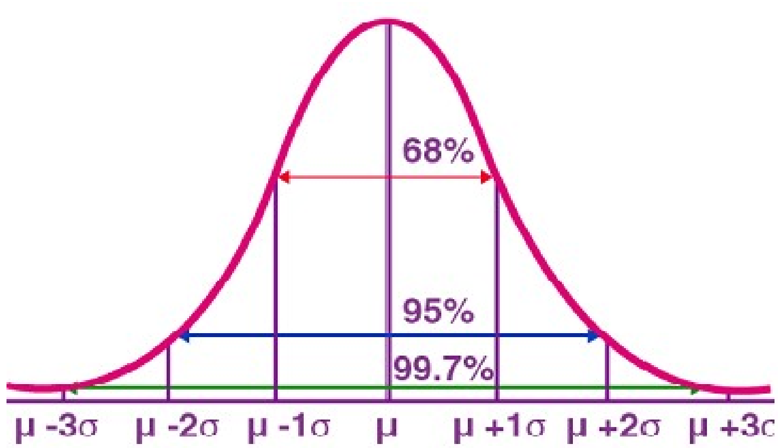


Fig:8.3.4

Ref:https://byjus.com/maths/normal-distribution/

The probability that a random value of X lies in the interval $(\mu - \sigma, \mu + \sigma)$ is given by:

$P(\mu-\sigma < X < \mu+\sigma) = \int_{\mu-\sigma}^{\mu+\sigma} f(x)dx \Rightarrow P\left(-1 < Z < 1 = \int_{-1}^{1} \varphi(z)dz\right) = 2 \times 0.3413 =$
0.6826…………..(8.3.4)
$P(\mu-2\sigma < X < \mu+2\sigma) = \int_{\mu-2\sigma}^{\mu+2\sigma} f(x)dx \Rightarrow$
$P\left(-2 < Z < 2 = \int_{-2}^{2} \varphi(z)dz\right) = 2 \times 0.4772 =$
0.9544…………..(8.3.5)
and

$$P(\mu-3\sigma < X < \mu+3\sigma) = \int_{\mu-3\sigma}^{\mu+3\sigma} f(x)dx \Rightarrow P\left(-3 < Z < 3 = \int_{-3}^{3} \varphi(z)dz\right) = 2\int_{0}^{3} \varphi(z)dz = 2 \times 0.49865 = 0.9973 \ldots\ldots\ldots\ldots (\mathbf{8.3.6})$$

Thus the probability that a normal variate X lies outside the range $\mu \pm 3\sigma$ is given by : $P(|X-\mu| > 3\sigma) = P(|Z| > 3) = 1 - P(-3 \leq Z \leq 3) = 0.0027$.

- The total area under normal probability curve is unity, *i.e.,*

$$\int_{-\infty}^{\infty} f(x)\varphi(z)dx = \int_{-\infty}^{\infty} \varphi(z)dz = 1.$$

In simple words the normal distribution has any positive standard deviation If the standard deviation is smaller, the data are somewhat close to each other and the graph becomes narrower. If the standard deviation is larger, the data are dispersed more, and the graph becomes wider. The standard deviations are used to subdivide the area under the normal curve. Each subdivided section defines the percentage of data, which falls into the specific region of a graph.

- Approximately 68% of the data falls within one standard deviation of the mean. (i.e., Between Mean- one Standard Deviation and Mean + one standard deviation).
- Approximately 95% of the data falls within two standard deviations of the mean. (i.e., Between Mean- two Standard Deviation and Mean + two standard deviations).
- Approximately 99.7% of the data fall within three standard deviations of the mean. (i.e., Between Mean- three Standard Deviation and Mean + three standard deviations)

Error Function:

The normal probability curve with mean μ and standard deviation σ is given by the equation *i.e.* if $X \sim N(\mu, \sigma^2), then\ f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}, -\infty < x < +\infty$

If we take $h^2 = \frac{1}{2\sigma^2} e^{-x^2/2\sigma^2}, f(x) = \frac{h}{\sqrt{\pi}} e^{-h^2x^2}$.

The probability P that a random value of the variate lies in the range $\pm x$is :

$$P = \int_{-X}^{X} f(x)dx = \frac{h}{\sqrt{\pi}} \int_{-x}^{x} e^{-h^2} x^2 dx$$

$$= \frac{2}{\sqrt{\pi}} \int_0^x e^{-h^2x^2} (hdx) \dots \dots .. (\mathbf{8.3.7})$$

Taking $\psi(y) = \frac{2}{\sqrt{x}} \int_0^y e^{-y^2} dy$, equation (8.3.7) may be re-written as:

$$P = \psi(hx)$$

$$= \frac{2}{\sqrt{\pi}} \int_0^x e^{-h^2x^2} (hdx) \dots \dots \dots \dots \dots \dots \dots \dots \dots (\mathbf{8.3.8})$$

The function $\psi(y)$, known a the error function.

8.3.5.IMPORTANCE OF NORMAL DISTRIBUTION:-

The normal distribution is an absolutely continuous distribution that plays a major role in statistics. Unlike the examples we have seen thus far, the normal distribution has a nonzero density function over the entire real number line. The normal distribution is determined by two parameters: the mean and the vari-ance. The fact that the mean and the variance of the normal distribution are the nat-ural parameters for the normal distribution explains why they are sometimes pre-ferred as measures of location and scale.For a normal distribution, there is no need to make the distinction among the mean, median, and mode. They are all equal to one another.

In the 1890s in England, Sir Francis Galton found applications for the normal distribution in medicine; he also generalized it to two dimensions as an aid in ex-plaining his theory of regression and correlation. In the 20th century, Pearson, Fisher, Snedecor, and Gosset, among others, further developed applications and other distributions including the chi-square, *F* distribution, and Student's *t* distribution, all of which are related to the normal distribution. Some of the most important early applications of the normal distribution were in the fields of agriculture, medicine, and genetics. Today, statistics and the normal distribution have a place in almost every scientific endeavor.

Although the normal distribution provides a good probability model for many phenomena in the real world, it does not apply universally. Other parametric and nonparametric statistical models also play an important role in medicine and the health sciences.

.

8.3.6.FITTING OF NORMAL DISTRIBUTION:-

To fit a normal curve to the observed data we first find the mean and variance from the given data. Mean and variance so obtained are μ and standard σ respectively. Then the normal curve fitted to the given data is given by:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}, -\infty < x < \infty$$

To calculate the expected normal frequencies we first find the standard normal variates corresponding to the lower limits of each of the class intevals, i.e., we compute $z_i = (x_i' - \mu)/\sigma$, where x_i' is the lower limit of the ith class interval. Then the areas under the normal curve to left of the ordinate at $z = z_i$, say $\Phi(z_i) = P(Z \leq z_i)$ are computed from the tables. Finally, the areas for the successive class intervals are obtained by subtraction, *viz.*, $\Phi(z_i + 1) - \Phi(z_i), (i = 1,2, \dots \dots ..)$ and on multiplying these areas by N, We get the expected normal frequencies.

8.4 .CENTRAL LIMIT THEOREM:-

The central limit theorem ranks high amongst the most important discoveries in the field of mathematics over the last three hundred years. This theorem provided a basis for approximation that turned the question of reaction into the art of prediction.The most basic form of the result is as follows; when we have a large number of independent random variables, the central limit theorem helps calculate how probable a certain deviation is away from the sum of said random variables in simple words "If a large random sample is taken from any distribution with mean μ and variance σ^2 , regardless of whether this distribution is discrete or continuous, then the distribution of the sample mean will tend to a normal distribution with mean $\mu$and variance σ^2/n."

If $X_i (i = 1,2, \dots\dots, n)$ be independent random variables such that $E(X_i) = \mu_i$ and $V(X_i) = \sigma_i^2$, then under certain very general conditions, the random variable $S_n = X_1 + X_2 + \cdots X_n$, is asymptotically normal with mean μ and standard deviation σ where

$$\mu = \sum_{i=a}^{n} \mu_i and\ \sigma^2 = \sum_{i=1}^{n} \sigma_i^2 \dots\dots\dots\dots\dots\dots\dots (\mathbf{8.4.1})$$

8.4.1. DE-MOIVRE'S LAPLACE THEOREM:-

A particular case of central limit theorem is De-Moivre's theorem which states as follows:

"If $X_i = \begin{cases} 1, with\ probability\ p \\ 0, with\ probability\ q \end{cases}$

Then the distribution of the random variable $S_n = X_1 + X_2 + .. + X_{n,}$ where $X_i's$ are independent, is asymptotically normal as $n \longrightarrow \infty$." ……………………***(8.4.2)***

Proof. Moment generating function of X_i is given by :

$M_{X_i}(t) = E(e^{tX}) = e^{t.1}p + e^{t.0} = (q + pe^t)$……………………………..(8.4.3)

M.G.F. of the sum $S_n = X_1 + X_2 + \cdots X_n$, is given by

$$M_{S_n}(t) = M_{X_1+X_2+\cdots X_n}(t) = M_{X_1}(t).M_{X_2}(t)\dots..M_{X_n}(t)$$
$$= \left[M_{X_1}(t)\right]^n$$

(Since $X_i's$ are independent and identically distributed.)

By using ***(8.4.3)*** $M_{S_n}(t) = [(q + pe^t)]^n$,

which is the moment generating function of a binomial variate with parameters n and p.

Hence by uniqueness theorem of $m.g, f's, S_n \sim B(n, p)$.

Therefore $E(S_n) = np = \mu, \& V(S_n) = npq = \sigma^2$,

Let $Z = \frac{S_n - E(S_n)}{\sqrt{V(S_n)}} = \frac{S_n - \mu}{\sigma}$.

$M_Z(t) = e^{-\mu(t/\sigma)} M_{S_n}(t/\sigma) = e^{-npt/\sqrt{npq}}\left[q + pe^t/\sqrt{npq}\right]^n$ ***[from (8.4.3)]***

$= \left[1 + \frac{t^2}{2n} + O\left(n^{-\frac{3}{2}}\right)^n\right]$,

Where $O\left(n^{-\frac{3}{2}}\right)$ represents terms involving $n^{3/2}$ and higher power of n in the denominator.

Proceeding to the limits as $n \longrightarrow \infty$, we get

$$\lim_{n\to\infty} M_Z(t)$$
$$= \lim_{n\to\infty}\left[1 + \frac{t^2}{2n} + O\left(n^{-3/2}\right)\right]^n$$
$$= \lim_{n\to\infty}\left[1 + \frac{t^2}{2n}\right]^n = e^{t^2/2},$$

Which is the moment generating function of a standard normal variate. Hence by uniqueness theorem of M.G.F's, $Z = \frac{S_n-\mu}{\sigma}$ is asymptotically $N(0,1)$. Hence $S_n = X_1 + X_2 + \cdots X_n$is asymptotically $N(\mu, \sigma^2)$ as $n \longrightarrow \infty$.

- Binomial distribution tends to normal distribution as $n \longrightarrow \infty$.

- Convergence in Distribution or Law. Let $\{X_n\}$ be sequence of random variables and $\{F_n\}$ be the corresponding sequence of distribution functions.
  It means that X_n converges in distribution to X if there exists a random variable X with distribution F such that as $n \longrightarrow \infty, F_n(x) \longrightarrow F(x)$ at every point x at which F is continuous.

8.5.SOLVED EXAMPLE:-

Example.8.5.1A sample of 100 dry battery cell tested to find the length of life produced the following results :

$\bar{x}$ = 12 hours, σ = 3 hours.

Assuming the data to be normally distributed, what percentage of battery cells are expected to have life

a) More than15 hours
b) less than 6 hours
c) between 10 and 14 hours.?

Solution. Here x denotes the length of life of dry battery cells.

Also $z = \frac{x-\bar{x}}{\sigma} = \frac{x-12}{3}$

a) When x = 15, z = 1

P(x>15) = P(z > 1)

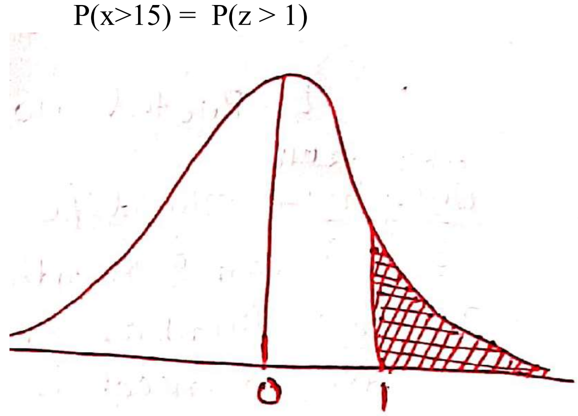


Fig.8.5.1

= P(0 < z < ∞) - P(0 < z < 1)

= 0.5 – 0.3413 = 0.1587 = 15.87%.

b) When x = 6, z = -2

P(x< 6) = P(z < -2)

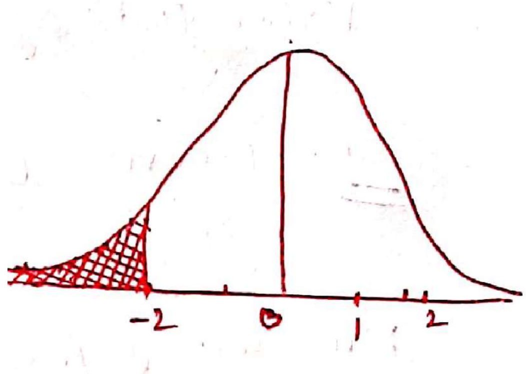


Fig.8.5.2

= p(z > 2) = P(0 < z < ∞) - P(0 < z < 2)
= 0.5 – 0.4772 = 0.0228 = 2.28%.

c) When x = 10, z = - $\frac{2}{3}$= -0.67

When x = 10, z = $\frac{2}{3}$= 0.67

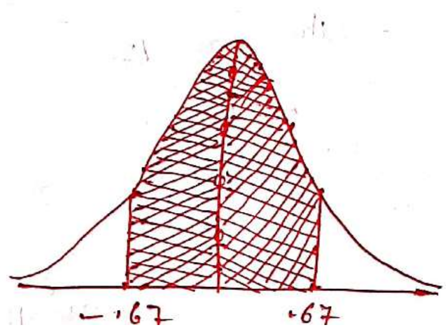


Fig.8.5.3

$P(10 < x < 14) = P(-0.67 < z < 0.67)$
$= 2\,P(0 < x < o.67) = 2 \times 0.2485 = 0.4970$
$= 49.70\%.$

Example.8.5.2 In a sample of 1000 cases, the mean of a certain test is 14 and S.D. is 2.5. Assuming the distribution to be normal, find

- i How many students score between 12 and 15 ?
- ii How many score above 18 ?
- iii How many score below 8 ?
- iv How many score 16 ?

Solution. (i) $z_1 = \frac{x_1 - \mu}{\sigma} = \frac{12-14}{2.5} = -0.8$

$z_2 = \frac{x_2 - \mu}{\sigma} = \frac{15-14}{2.5} = 0.4$

Area lying between -0.8 and 0.4

= area between 0 to 0.8 + area between 0 to 0.4
= 0.2881 + 0.1554 = 0.4435

Required no. of students = 1000 × 0.4435 = 444 (app.)

(ii) $z = \frac{18-14}{2.5} = 1.6$

Area right to 1.6 = 0.5 – (area between 0 and 1.6) = 0.5 – 0.4452 = 0.0548

Required no. of students = 1000 × 0.0548 = 54.8 = 55 (app.)

(iii) $z = \frac{8-14}{2.5} = -2.4$

Area left to -2.4 = 0.5 – (area between 0 and 2.4) = 0.5 – 0.4918 = 0.0082

Required no. of students = 1000 × 0.0082 = 8.2 = 8 (app.)

(iv) $z_1 = \frac{x_1 - \mu}{\sigma} = \frac{15.5-14}{2.5} = 0.6$

$z_1 = \frac{x_2 - \mu}{\sigma} = \frac{16.5-14}{2.5} = 1$

Area between 0.6 and 1 = 0.3413 – 0.2257 = 0.1156

Required no. of students = 1000 × 0.1156 = 115.6 = 116 (app.)

Example8.5.3 Assume mean height of soldiers to be 68.22 inches with a variance of 10.8 inches with a variance of 10.8 inches square. How many soldiers in a regiment of 1,000 would you expect to be over 6 feet tall, given that the area under the standard normal curve between z = 0 and z = 0.35 is 0.1368 and between z = 0 and z = 1.15 is 0.3746.

Solution. X = 6 feets = 72 inches

$$Z = \frac{x - \mu}{\sigma} = \frac{73-68.22}{\sqrt{10.8}} = 1.15$$

P(x> 72) = P(x > 1.15) = p($0 \le z \le 1.15$)
= 0.5 – 0.3746 = 0.1254

Expected no. of soldiers = 1000 × 0.1254 = 125.4 = 125 (app.)

Example8.5.4A largenumber of measurement is normally distributed with a mean 65.5'' and S.D. of 6.2''. find the percentage of measurements that fall between 54.8'' and 68.8''.

Solution. Mean μ = 65.5 inches, S.D. σ 6.2 inches

x_1 = 54.8 inches, x_2 = 68.8 inchesf

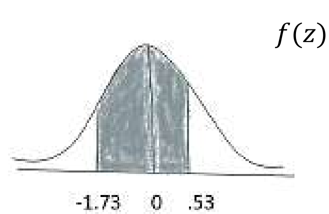


Fig.8.5.4

$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{54.8 - 65.5}{6.2} = -1.73$ and $z_2 = \frac{x_2 - \mu}{\sigma} = \frac{68.8 - 6\ .5}{6.2} =$ 0.53 now , P(-1.73 ≤ x ≤ 0.53) = P(-1.73 ≤ x ≤ 0) + P(0 ≤ x ≤ 0.53).

= P(0 ≤ x ≤ -1.73) + P(0 ≤ x ≤ 0.53)

= 0.4582 + 0.2019 = 0.6601 (By table)

Required percentage of measurements = 66.01%.

Example8.5.5 A manufacturer knows from experience that the resistance of resistors he produces is normal with mean μ = 100 ohms and standard deviation σ = 2 ohms. What percentage of resistors will have resistance between 98 ohms and 102 ohms?

Solution.μ = 100 Ω, σ = 2 Ω $x_1 = 98$ Ω $x_2 = 102$ Ω

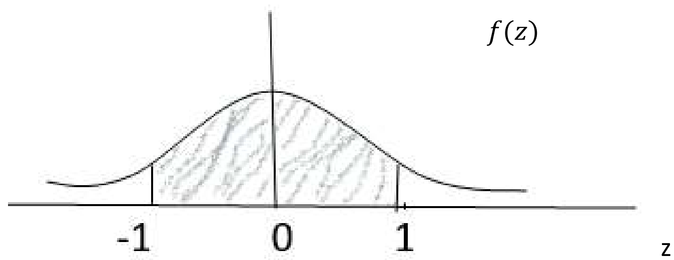


Fig.8.5.5

$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{98 - 100}{2} = -1$$

and $z_2 = \frac{x_2 - \mu}{\sigma} = \frac{102 - 100}{2} = 1$

Now, P(98 < x < 102) = P(-1 < z < 1)

= P(-1 ≤ z ≤ 0) + P(0 ≤ z ≤ 1)

= P(0 ≤ z ≤ 1) + P(0 ≤ z ≤ 1)

= 0.3413 + 0.3413 = 0. 6826

Percentage of resistors having resistance between 98 Ω and 102 Ω = 68.26%.

Example 8.5.6. In a normal distribution, 31% of the items are under 45% and 8% are over 64.Find the mean and standard deviation of the distribution. It is given that if $f(t) = \frac{1}{\sqrt{2\pi}} \int_0^t e^{-\frac{x^2}{2}}\, dx$ then f(0.5) = 0.19 and f (1.4) = 0.42.

Solution. Let μ and σ be the mean and S.D. respectively.

31% of the items are under 45.

⇒ Area to the left of the ordinate x = 45 is 0.31

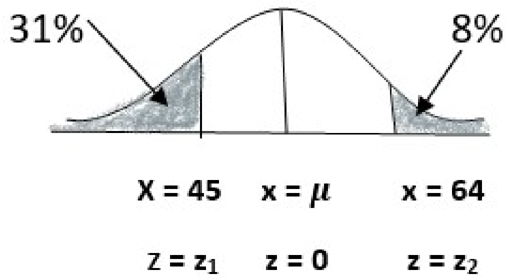


Fig.8.5.6

When x = 45, let $z = z_1$
$P(z_1 < z < 0) = 0.5 - 0.31 = 0.19$
From the tables, the value of z corresponding to this area is 0.5
$\therefore\ z_1 = -0.5\ [z_1 < 0]$
When x = 64, let $z = z_2$
$P(z_1 < z < 0) = 0.5 - 0.08 = 0.42$
From the tables, the value of z corresponding to this area is 1.4.
$\therefore\ z_2 = 1.4$
Since $z = = \frac{x-\mu}{\sigma}$
$-0.5 = = \frac{45-\mu}{\sigma}$ and $1.4 = = \frac{64-\mu}{\sigma}$
$\Rightarrow\ 45 - \mu = -0.5\sigma$ ....(1)
and
$64 - \mu = 1.4\sigma$ ....(2)
Subtracting $-19 = -1.9\sigma$ $\therefore \sigma = 10$
From (1), $45 - \mu = -0.5 \times 10 = -5$ $\therefore \mu = 50$.

Example8.5.7 The life of army shoes is normally distributed with mean 8 months and standard deviation 2 months. If 5000 pairs are insured, how many pairs would be expected to need replacement after 12 months? [Given that $P(z \geq 2) = 0.0228$ and $z = = \frac{x-\mu}{\sigma}$].

Solution. Mean (μ) = 8, S.D. (σ) = 2
Number of pairs of shoes = 5000, Total months (x) = 12
When x = 12, $z = \frac{x-\mu}{\sigma} = \frac{12-8}{2} = 2$
Area $(z \geq 2) = 0.0228$
Number of pairs whose life is more than 12 months = 5000 – 114 = 4886.

Example8.5.8 The mean inside diameter of a sample of 200 washers produced by a machine is 0.502 cm and standard deviation is 0.005 cm. the purpose for which these washers are intended allows a minimum tolerance in the diameter of 0.496 to 0.508 cm, otherwise the washers are considered defective. Determine the percentage of defective washers produced by the machine. Assume the diameters are normally distributed.

Solution. Given, Mean $\mu = 0.502$ cm, S.D. $\sigma = 0.005$ cm, $x_1 = 0.496$ cm, $x_2 = 0.508$.

Now, $z_1 = \frac{x_1 - \mu}{\sigma} = \frac{0.496 - 0.502}{0.005} = -1.2$

and $z_2 = \frac{x_2 - \mu}{\sigma} = \frac{0.508 - 0.502}{0.005} = 1.2$

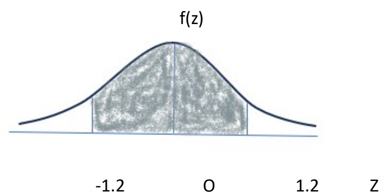


Fig.8.5.7

Area for non-defective washers

$= P(-1.2 \le z \le 1.2)$
$= P(-1.2 \le z \le 0) + P(0 \le z \le 1.2)$
$= P(0 \le z \le 1.2) + P(0 \le z \le 1.2)$
$= 0.3849 + 0.3849$
$= 0.7698$
$= 76.98\%$.

$\therefore$ Percentage of defective washers = 100 – 76.98
= 23.02%.

Example8.5.9 Assuming that the diameters of 1000 brass plugs taken consecutively from a machine, from a normal distribution with mean 0.7515 cm and standard deviation 0.002 cm, how many of the plugs are likely to be rejected if the approved diameter is 0.752 ± 0.004 cm.

Solution. Tolerance limits of the diameter of non-defective plugs are

0.752 – 0.004 = 0.748 cm.and 0.752 + 0.004 = 0.756 cm.

Standard normal variable, $z = \frac{x - \mu}{\sigma}$

If $x_1 = 0.748$, $z_1 = \frac{0.748 - -0.7515}{0.002} = -1.75$

If $x_2 = 0.756$, $z_2 = \frac{0.756 - 0.7515}{0.002} = 2.25$

Area from ($z_1 = -1.75$) to ($z_2 = 2.25$)

$= P(-1.75 \le z \le 2.25) = P(-1.75 \le z \le 0) + P(0 \le z \le 2.25)$
$= P(0 \le z \le 1.75) + P(0 \le z \le 2.25)$
$= 0.4599 + 0.4878 = 0.9477$

Number of plugs which are likely to be rejected = 1000 × (1 – 0.9477) = 1000 × 0.523 = 52.3

Hence approximately 52 plugs are likely to be rejected.

Example.8.5.10 If the height of 300 students are normally distributed with mean 64.5 inches and Standard deviation 3.3 inches, find the height below which 99% of the students lie.

Solution. Mean $\mu = 0.502$ inches, S.D. $\sigma = 0.005$ inches

Area between 0 and $\frac{x-64.5}{3.3} = 0.99 - 0.5 = 0.49$

From the table, for the area 0.49, z = 2.327

The corresponding value of x is given by

$\frac{x-64.4}{3.3} = 2.327$

$\Rightarrow \quad x - 64.5 = 7.68$

$\Rightarrow \quad x = 7.68 + 64.5 = 72.18$ inches.

Hence 99% students are of height less than 6 ft. 0.18 inches.

Example.8.5.11 The income of a group of 10,000 persons was found to be normally distributed with mean Rs. 750 and standard deviation of Rs. 50. Show that, of this group, about 95% had income exceeding Rs. 668 and only 5% had income exceeding Rs. 832. Also find the lowest income among the richest 100.

Solution. Given $\mu = 750$, $\sigma = 50$

Standard normal variable, $z = \frac{x-\mu}{\sigma}$

i If $x_1 = 668$, $z_1 = \frac{x_1-\mu}{\sigma} = \frac{668-750}{50} = -1.64$

$P(x_1 > 668) = P(z_1 > -1.64)$
$= 0.5 + P(-1.64 \le z \le 0)$
$= 0.5 + P(0 \le z \le 1.64)$
$= 0.5 + 0.4495$
$= 0.9495$

∴ Required percentage of persons having income exceeding Rs. 668 = 94.95% = 95% (approx.)

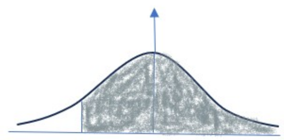


Fig 8.5.11

ii. If $x_1 = 832$, $z_1 = \frac{x_2 - \mu}{\sigma} = \frac{8 - 750}{50} = -1.64$

$P(x_1 > 832) = P(z_2 > 1.64)$

$= 0.5 - P(0 \leq z \leq 1.64)$

$= 0.5 - 0.4495$

$= 0.0505$

∴ Required percentage of persons having income exceeding Rs. 832 = 5.05% = 5% (approx.)

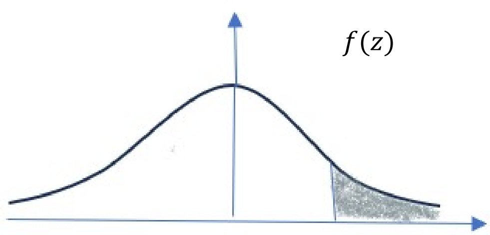


Fig 8.5.12

iii. let x be the lowest income among the richest 100 persons i.e., 1% of 10,000.Thus, area between o and z = 0.49 (see figure) by Normal distribution table,

$Z = 2.33$

Thus, $\frac{x - \mu}{\sigma} = 2.33$

$\Rightarrow \frac{x - 750}{50} = 2.33$

$\Rightarrow x = 866.5$

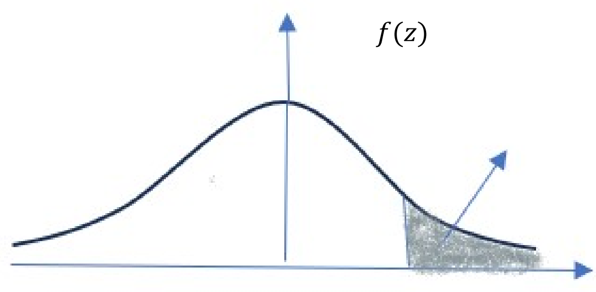


Fig 8.5.13

Hence Rs. 866.5 is the minimum income among the richest 100 persons.

Example.8.5.12 255 mental rods were cut roughly 6 inches over size. Finally the lengths of the over size amount, were measured exactly and grouped with 1 inch intervals, there being in all 12 groups $\frac{1}{2}$ ‘’, -1$\frac{1}{2}$ ‘’ , 1$\frac{1}{2}$ ‘’, -2$\frac{1}{2}$ ‘’, …… , 11$\frac{1}{2}$ ″, -12$\frac{1}{2}$’’.

The frequency distribution for the 255 lengths was as follows:

Length(inches) Central value	1	2	3	4	5	6	7	8	9	10	11	12
Frequency	2	10	19	25	40	44	41	28	25	15	5	1

Fit a normal curve to this data.

Solution. The equation of the normal curve for N observation is

x	**f**	**u = x – 6**	**fu**	**fu²**
1	2	-5	-10	50
2	10	-4	-40	160
3	19	-3	-57	171
4	25	-2	-50	100
5	40	-1	-40	40
6	44	0	0	0
7	41	1	41	41
8	28	2	56	112

9	25	3	75	225
10	15	4	60	240
11	5	5	25	125
12	1	6	6	36
Total	255		66	1300

Mean, $\mu = \alpha + \frac{\Sigma fu}{\Sigma f} = 6 + \frac{66}{255} = 6.259$

Variance, $\sigma^2 = \frac{\Sigma fu^2}{\Sigma f} = \left(\frac{\Sigma fu}{\Sigma f}\right)^2 = \frac{1300}{225} - \left(\frac{66}{255}\right)^2 = 5.031$

$\therefore \sigma = 2.243$

Thus, we have N = 255, Mean, μ = 6.259'' , S.D. σ = 2.243''

Hence the fitted curve is

$$Y = \frac{255}{2.243\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-6.259}{2.243}\right)^2} \quad \text{From (1)}$$

$$= \frac{113.68}{\sqrt{2\pi}} e^{-0.009(x-6.259)^2}.$$

Example.8.5.13 Show that the area under the normal curve is unity.

Solution. Area under the normal curve is given by

$$A = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

Put $\frac{x-\mu}{\sigma} = z$ so $dx = \sigma\, dz$

$$\therefore A = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{z^2}{2}} (\sigma dz) = \frac{\sqrt{2}}{\sqrt{\pi}} \int_0^{\infty} e^{-\frac{z^2}{2}} dz$$

Now, $A \cdot A = A^2 = \left(\frac{\sqrt{2}}{\sqrt{\pi}} \int_0^{\infty} e^{-\frac{x^2}{2}} dz\right)\left(\frac{\sqrt{2}}{\sqrt{\pi}} \int_0^{\infty} e^{-\frac{x^2}{2}} dz\right)$

$= \frac{2}{\pi} \int_0^{\infty} \int_0^{\infty} e^{-\left(\frac{x^2+y^2}{2}\right)} dx\, dy$ (where x and y are dummy variables)

Put $x = r\cos\theta$, $y = r\sin\theta$ so that J = r changing to polar coordinates,

$$A^2 = \frac{2}{\pi} \int_0^{\pi/2} \int_0^{\infty} e^{-\frac{r^2}{2}} r\, dr\, d\theta = \int_0^{\infty} e^{-\frac{r^2}{2}} d\left(\frac{r^2}{2}\right) = 1$$

$$\therefore A = Area\ under\ the\ normal\ curve = 1$$

Example.8.5.14 Prove that for normal distribution, the mean deviation from the mean equals to $\frac{4}{5}$ of the standard deviation approximately.

Solution. Let μ *and* σ *be* the mean and standard deviation of the normal distribution. Then by definition,

Mean deviation from the mean

$$= \int_{-\infty}^{\infty} |x - \mu|\, f(x)\, dx$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} |x - \mu|\, e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \sigma |\, z \,| \, e^{-\frac{1}{2}z^2} \sigma \, dz$$

$$= \sigma \sqrt{\frac{2}{\pi}} \int_{0}^{\infty} \sigma \; e^{-\frac{1}{2}z^2} \, dz$$

$$= \sigma \sqrt{\frac{2}{\pi}} [\; -e^{-\frac{1}{2}z^2} \;] = \sqrt{\frac{2}{\pi}} \, \sigma = 0.7979 \; \sigma \; = \; 0.8$$

$$\sigma = \frac{4}{5}\sigma \; .$$

CHECK YOUR PROGRESS

Fill in the blank

Q.1 The normal distribution is ………. Distribution.

Q.2 The normal distribution is symmetrical about its ……….

Q.3 The mean, mode and median of the normal distribution are ……….

True\false

Q.4 The total area under the normal curve above the x – axis is 1.

Q.5 The probability density function for the normal distribution in standard form is $f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$.

Q.6 If Z is a standard normal variable, probabilities $P(z < 1.2)$ is 0.8849.

Q.7 The graph of the normal distribution is called the linear curve.

Objective questions

Q.8 The marks X obtained in Mathematics by 1000 students are normally distributed with mean 78% and standard deviation 11%, then how many students got marks above 90%

(a) 128 (b) 138 (c) 148 (d) None

Q.9 De Moivre made the discovery of the normal distribution in year

(a) 1733 (b) 1800 (c) 1900 (d) 2000

Q.10 The graph of normal distribution is bell-shaped and symmetrical about the

(a) mean μ (b) variance σ (c) S.D. (d) y – axis

8.6.SUMMARY:-

Present unit is an explanation of continuous probability distribution andcentral limit theorem. In continuous probability distribution our main target is Normal distribution. In normal distribution we have discussed aboutLimiting Form of Binomial Distribution, Normal Probability Curve, Properties of Normal Distribution, Area property, Importance of Normal Distribution and Fitting of Normal Distribution. In central limit theorem we have discussed about De-Moivre's Laplace Theorem.

8.7. GLOSSARY:-

- i Mean
- ii Variance
- iii Probability Density Function
- iv Moments
- v Mathematical expectation
- vi Moment generating function.
- vii Random variable

8.8.REFERENCE BOOKS:-

1. S. C. Gupta and V. K. Kapoor, (2020), *Fundamentals of mathematical statistics*, Sultan Chand & Sons.
2. Seymour Lipschutz and John J. Schiller, (2017), *Schaum's Outline: Introduction to Probability and Statistics*, McGraw Hill Professional.
3. J. S. Milton and J. C. Arnold , (2003*), Introduction to Probability and Statistics (4th Edition),* Tata McGraw-Hill.
4. https://www.wikipedia.org.

8.9. SUGGESTED READINGS:-

1. A.M. Goon,(1998), *Fundamental of Statistics 7th Edition),* 1998.
2. R.V. Hogg and A.T. Craig, (2002), *Introduction to Mathematical Statistics,* MacMacMillan, 2002.
3. Jim Pitman, (1993), *Probability,* Springer-Verlag.
4. https://archive.nptel.ac.in/courses/111/105/111105090.

8.10. TERMINAL QUESTIONS :-

TQ1. The mean yield for one – acre plot is 662 $kilos$ with a$s.d.$ 32 $kilos$. Assuming normal distribution, how many one-acre plots in a batch of 1,000 plots would you except to have yield $(i) over$ 700 $kilos$, $(ii) below$ 650 $kilos$, and (iii)what is the lowest yield of the best 100 plots?

TQ2. There are six hundred Economics learners in the post-graduate classes of a university, and the probability for any learner to need a copy of a particular book from the university library on any day is 0.05. How many copies of the book should be kept in the university library so that the probability may be greater than 0.90 that none of the learners needing a copy from the library has to come back disappointed ? (Use nomal approximation to the binomial distribution?

TQ3.(i) If $\log_{10} X$ is normally distributed with meant 4 and variance 4, find the probability of $1.202 < X < 83180000$.(Given $\log_{10} 1202 = 3.08, \log_{10} 8318 = 3.92$).
(ii)$\log_{10} X$is normally distributed with mean7and variance 3, $\log_{10} Y$ is normally distributed with 3 and variance unity. If the distributions of X and Y are independent, find the probability of $1.202 < (X/Y) < 83180000$.(Given $\log_{10} 1202 = 3.08, \log_{10} 8318 = 3.92$).

TQ4. In a distribution exactly normal, 10.03% of the items are under 25 Kilogram weight and 89.97% of the items are under 70 kilogram weight. What are the mean and standard deviation of the distribution?

8.11.ANSWERS:-

Answer of Check your progress Questions:-

CHQ.1 Continuous
CHQ.2 Mean
CHQ.3 Coincide
CHQ.4 True
CHQ.5 True
CHQ.6 True
CHQ.7 False
CHQ.8 (b)
CHQ.9(a)
CHQ.10(a)

Answer of Terminal Questions:-

TQ1: (i)0.1170 (ii)352 (iii)702.96.
TQ2: 37 copies of the book.
TQ3: (i) Required probability = 0.9500 (ii) Required probability = 0.95
TQ4: Mean is47.5 kilogram and standard deviation is 17.578 kilogram

BLOCK -III
CORRELATION& REGRESSION

UNIT 9:-CORRELATION

CONTENTS:

9.1. INTRODUCTION:-

Till now we have studied various statistical measure for univariate data. In this unit we will study the bivariate data. Suppose for a bivariate data, we want to search a connection between given data, then we try to see these data are correlated are not. In statistics there is mathematical computative technique, which serve our purpose. This technique is known as correlation. In this unit we have defined about Bivariate distribution and Correlation, Correlation coefficient, Correlation Coefficient for a bivariate frequency distribution and Rank Correlation.

9.2. OBJECTIVES:-

After studying this unit learner will be able to:

1. Analyse the correlation between two variables.
2. Compute the correlation coefficient.
3. Evaluate the rank correlation.

9.3. BIVARIATE DISTRIBUTION & CORRELATION:-

In a bivariate distribution we may be interested to find out if there is any correlation or covariation between the two variables under study. If the change in one variable affects a change in the other variable, the variables are said to be correlated. If the two variables deviate in the same direction, $i.e.,$ after the increasing or decreasing the value of one variable gives increase or decrease respectively in the value of second variable, then correlation is said to be direct or positive. But if they constantly deviate in the opposite directions, $i.e.,$ if increase (or decrease) in one variable results in corresponding decrease (or increase) in the other variable, correlation is said to be diverse of negative.

Example Consider the following bivariate conditions:

- **i.** the heights and weights of a group of persons
- **ii.** the income and expenditure
- **iii.** price and demand of a commodity
- **iv.** the volume and pressure of a perfect gas

In above four scenario, $(\boldsymbol{i})$ and $(\boldsymbol{ii})$ are positive correlated and $(\boldsymbol{iii})$ & $(\boldsymbol{iv})$ are negatively correlated.

Scatter Diagram. Consider a bivariate distribution $(x_i, y_i);\ i = 1, 2, \ldots, n$. Its scatter diagram is representing these points in $X - Y$ plane. It is the simplest way of the diagrammatic representation of bivariate data. The values of the variables X and Y be plotted along the x-axis and y-axis respectively. After that we mark each entries$(x_i, y_i);\ i = 1, 2, \ldots, n$ in this $X - Y$ plane. The diagram of dots so obtained is known as scatter diagram. From the scatter diagram, we can form a fairly good, idea whether the variables are correlated or not, e.g..if the points are very dense $i.e.,$ very close to each other, we should expect a fairly good amount of correlation between the variables and if the points are widely scattered, a poor correlation is expected. This method however is not suitable if the number of observations is fairly large.

9.4.CORRELATION COEFFICIENT:-

KARL PEARSON'S COEFFICIENT OF CORRELATION. To measure the correlation between the variables in Karl Pearson (1867-1936) gives a formula called Correlation Coefficient.

Correlation coefficient between two random variables $X$and Y, usually denoted by $r(X\,Y)$ or simplyr_{xy}, is a numerical measure of linear relationship between them and is defined as

$$r(X,Y)=\frac{\text{Cov}(X,Y)}{\sigma_x\sigma_y}$$

If $(x_i,y_i);\ i\ =\ 1,2,\ldots,n$ is the bivariate distribution, then

$$\text{Cov}(X,Y)=E[\{X-E(X)\}\ \{(Y-E(Y)\}]$$
$$=\frac{1}{n}\ \sum(x_i-\bar{x})(y_i-\bar{y})=\mu_{11}$$
$$\sigma_x{}^2=E\{X-E(X)\}^2=\frac{1}{n}\ \sum(x_i-\bar{x})^2$$
$$\sigma_y{}^2=E\{Y-E(Y)\}^2=\frac{1}{n}\ \sum(y_i-\bar{y})^2$$

The summation extending over i from 1 to n. Another convenient form for computational work is as follows: $\text{Cov}(X,Y)=\frac{1}{n}\sum(x_i-\bar{x})(y_i-\bar{y})=\frac{1}{n}(x_iy_i-x_i\bar{y}-\bar{x}y_i+\bar{x}\bar{y})=\frac{1}{n}\sum x_iy_i-\bar{y}\frac{1}{n}\sum x_i-\bar{x}\frac{1}{n}\sum y_i+\bar{x}\bar{y}$
$\text{Cov}(X,Y)=\frac{1}{n}x_iy_i-\bar{x}\bar{y},\sigma_x{}^2=\frac{1}{n}\sum x_i{}^2-\bar{x}^2$ and $\sigma_y{}^2=\frac{1}{n}\sum y_i{}^2\bar{y}^2$

REMARKS 9.4.1

(i) It may be noted that $r(X,Y\,)$ provides a measure of *linear relationship* between X and Y.For nonlinear relationship, however, it is not very suitable.
(ii) Sometimes, we write $\text{Cov}(X,Y)=\sigma_x\sigma_y$.
(iii) Karl Pearson's correlation coefficient is also called product moment correlation coefficient, since. $\text{Cov}(X,Y)=E[\{X-E(X)\}\{(Y-E(Y)\}]=\mu_{11}$.

LIMITS FOR CORRELATION COEFFICIENT.

We know that

$$r(X,Y)=\frac{\text{Cov}(X,Y)}{\sigma_x\sigma_y}=\frac{\frac{1}{n}\sum(x_i-\bar{x})(y_i-\bar{y})}{\left[\frac{1}{n}\sum(x_i-\bar{x})^2.\ \frac{1}{n}\sum(y_i-\bar{y})^2\right]^{1/2}}$$

Squaring on both sides

$$\therefore\qquad r^2(X,Y)=\frac{(\sum a_ib_i)^2}{(\sum a_i)^2(\sum b_i)^2}\ ,\qquad \text{where}\ \begin{pmatrix}a_i=x_i-\bar{x}\\ b_i=y_i-\bar{y}\end{pmatrix}$$

In mathematics there is a Schwartz inequality which states that if $a_i,b_i\,;i=1,2,\ldots n$ are real quantities then

$$\left(\sum_{i=1}^{n}a_ib_i\right)^2\le\left(\sum_{i=1}^{n}a_i{}^2\right)\left(\sum_{i=1}^{n}b_i{}^2\right)$$

The sign of equality holding if and only if

$$\frac{a_1}{b_1}=\frac{a_2}{b_2}=\cdots=\frac{a_n}{b_n}$$

Using Schwartz inequality, we get $r^2(X,Y)\le1\ \ i.e.,|r(X,Y)|\le1\Rightarrow\ -1\le r(X,Y)\le1$.

Hence correlation coefficient cannot exceed unit numerical. It always lies between -1 and +1. If $r = +1$, the correlation is perect and positive and if $r = -1$, correlation is perfect and negative.

Theorem 9.4.1. Correlation coefficient is independent of change of origin and scale.

Proof. Let $U = \frac{X-a}{h}, V = \frac{Y-b}{k}$, so that $X = a + hU$ and $Y = b = kV$. Where a, b, h, k are constants; $h > 0, k > 0$. We shall prove that $r(X,Y) = r(U,V)$. Since $X = a + hU$ and $Y = b + kV$, on taking expectations, we get $E(X) = a + hE(U)$ and $E(Y) = b + kE(V)$. Therefore $X - E(X) = h[U - E(U)]$ and $Y - E(Y) = k[V - E(V)]$

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

$$= E[h\{U - E(U)\}\ \{k\{V - E(V)\}]$$
$$= hk\ E\ [\{U - E(U)\}\{V - E(V)\}]$$
$$= hk\ \text{Cov}(U,V) \qquad \dots\dots\dots\dots(\mathbf{9.4.1})$$

and $\sigma_x{}^2 = E[\{X - E(X)\}^2] = E[h^2\{U - E(U)\}^2] = h^2\sigma_U{}^2$

This implies that, $\sigma_x = h\sigma_U$, $(h > 0)$ ....$(\mathbf{9.4.2})$

Similarly,

$\sigma_Y{}^2 = E[\{Y - E(Y)\}^2] = E[k^2\{V - E(V)\}^2] = h^2\sigma_V{}^2$

$$\sigma_y = h\sigma_y\ , \qquad (k > 0) \quad \dots\dots(\mathbf{9.4.3})$$

Substituting equations **(9.4.1), (9.4.2) & (9.4.3)** in formula of $r(X,Y)$, we get $r(X,Y) = \frac{\text{Cov}\,(X,Y)}{\sigma_x\sigma_y} = \frac{hk.\ \text{Cov}(U,V)}{hk.\sigma_U\sigma_V} = \frac{\text{Cov}(U,V)}{\sigma_U\sigma_V} = r(U,V)$

This theorem is of fundamental importance in the numerical computation of the correlation coefficient.

Corollary. If X and Y are random variables and a, b, c, d are any numbers provided only that $a \neq 0, c \neq 0$, then

$$r(aX + b, cY + d) = \frac{ac}{|\,ac\,|}r\ (X,Y)$$

Proof. With usual notations, we have

$$\text{Var}(aX + b) = a^2\sigma_x{}^2\ ;\quad \text{Var}(cY + d) = c^2\sigma_y{}^2$$
$$\text{Cov}(aX + b, cY + d) = ac\sigma_{xy}$$

Therefore, $r\ (aX + b,\ cY + d) = \frac{\text{Cov}(aX+b,cY+d)}{[\text{Var}(aX+b)\text{Var}(cY+d)]^{\frac{1}{2}}}$

$$= \frac{ac\sigma_{xy}}{|a||c|\sigma_x\sigma_y} = \frac{ac}{|\,ac\,|}r(X,Y).$$

Note that, if $ac > 0$, $i.e.$, if a and c are of same signs, then $\frac{ac}{|\,ac\,|}$=+1. And if $ac < 0$, $i.e.$, if a and c are of opposite signs, then $\frac{ac}{|\,ac\,|}$=-1.

Theorem 9.4.2. Two independent variables are uncorrelated.

Proof. If X and Y are independent variables, then

$$\text{Cov}(X,Y) = 0$$

$$r(X,Y) = \frac{\text{Cov }(X,Y)}{\sigma_x \sigma_y} = 0$$

Hence two independent variables are uncorrelated. But the converse of the theorem is not true. Consider the following problem.

Problem 9.4.1: Give an example of, two uncorrelated variables which are not independent.
Solution: Consider the following data:

							Total
X	-3	-2	-1	1	2	3	$\sum X = 0$
Y	9	4	1	1	4	9	$\sum Y = 28$
XY	-27	-8	-1	1	8	27	$\sum XY = 0$

Table: 9.4.1

$$\bar{X} = \frac{1}{n}\Sigma X = 0, \ \ Cov(X,Y) = \frac{1}{n}\Sigma XY - \bar{X}\bar{Y} = 0$$

$$r(X,Y) = \frac{\text{Cov}(X,Y)}{\sigma_x \sigma_y} = 0$$

Thus, in the above example, the variables X and Y are uncorrelated. But on careful examination we find that X and Y are not independent but they are connected by the relation $Y = X$. Hence two uncorrelated variables need not necessarily be independent. A simple reasoning for this strange conclusion is that $r(X\,,Y) = 0$, merely implies the absence of any linear relationship between the variables X and Y. There may, however, exist some other form of relationship between them, $e.g.$, quadratic, cubic or trigonometric.

Remarks.9.4.2. .**(i)** Following are some more examples where two variables are uncorrelated but *not* independent.

($\boldsymbol{i}$) $X \sim N\,(0{,}1)$and $Y = X^2$.Since $X \sim N(0{,}1)$, $E(X) = 0 = E(X^3)$.Therefore,

$$\text{Cov}(X,Y) = E(X,Y) - E(X)\,E(Y) \ = E(X^3) - E(X)E(Y) = 0(\because Y = X^2).$$

This implies that, $r(X,Y) = \frac{\text{Cov}(X,Y)}{\sigma_x \sigma_y} = 0$.

Hence X and Y are uncorrelated but not independent.

(**ib**)Let X be a random variable with p.d.f.

$$f(x) = \frac{1}{2}, -1 \leq x \leq 1$$

and let $Y = X^2$. Then after computations $E(X) = 0$ and $E(XY) = E(X^3) = 0$. Consequently, $r(X,Y) = 0$

(ii) However, the converse of the theorem holds in the following cases:

(iia) If X and Y are jointly normally distributed with $\rho = \rho(X,Y) = 0$, then they are independent. If $\rho = 0$, then

$$f(x,y) = \frac{1}{\sigma_x\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{X-\mu_x}{\sigma_x}\right)^2\right] \times \frac{1}{\sigma_x\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{Y-\mu_y}{\sigma_y}\right)^2\right]$$

$f(x,y) = f_1(x)\, f_2(y)$ i.e. X and Y are independent.

(iib) If each of the two variables X and Y takes two values, 0,1 with positive probabities, then $\mathrm{r(X,Y)} = 0$, then X and Y are independent.

Proof (iib). Let X take the values 1 and 0 with positive probabilities p_1 and q_1 respectively and let Y take the values 1 and 0 with positive probabilities p_2 and q_2 respectively. Then $r(X,Y) = 0$ .This gives, $\mathrm{Cov}(X,Y) = 0$. And hence, $0 = E(XY) - E(X)E(Y) = 1$.
$P(\{X=1\} \cap \{Y=1\}) - [1\,.P(X=1)].[1.P(Y=1)] =$
$P(\{X=1\} \cap \{Y=1\}) - p_1p_2$
.i.e. $P(X=1 \cap Y=1) = p_1p_2 = P(X=1).(Y=1)$.Thus, X and Y are independent.

9.5.CORRELATION COEFFICIENT FOR A BIVARIATEFREQUENCY DISTRIBUTION:-

When the date are considerably large, they may be summarized by using a two-way table. Here, for each variable a suitable number of classes are taken, keeping in view the same considerations as in the univariate case. If there are n classes for X and m classes for Y, there will be in all $m \times n$ cells in the two-way table. By going through the pairs of values of and Y , we can find the frequency for each cell. The whole set of cell frequencies will then define a bivariate frequency distribution. The column totals and row totals will give us the marginal distributions of X and Y. A particular column or row will be called the conditional distribution of Y for given X or of X for given Y respectively. Suppose that the bivariate data on X and Y are presented in two-way correlation table where there are m classes of Y placed along the horizontal line and n classes of X along a vertival line and is frequency of individuals lying in the (i,j)th cell here $\sum_x f(x,y) = g(y)$ is the sum of the frequencies along any row and $\sum_y f(x,y) = g(x)\, is$ the sum of the frequencies along any column. We observe that

$$\sum_x\sum_y f(x,y) = \sum_x\sum_y f(x,y) = \sum_x f(x) = \sum_x g(y) = N$$

Thus

$$\bar{x} = \frac{1}{\mathrm{N}}\sum_x\sum_y x\, f(x,y) = \frac{1}{N}\left[\sum_x f(x)\sum_y f(x,y)\right] = \frac{1}{\mathrm{N}}\sum_x xf(x)$$

Similarly

$$\bar{y} = \frac{1}{N}\sum_x\sum_y y\, f(x,y) = \frac{1}{N}\sum_y y\, g(y)$$

$$\sigma_x{}^2 = \frac{1}{N}\sum_x\sum_y x^2 f(x,y) - \overline{x^2} = \frac{1}{N}\sum_x x^2 f(x) - \overline{x^2}$$

BIVARIATE FREQUENCY TABLE (CORRELATION TABLE)

<table>
<tr><td rowspan="2" colspan="2">X series →
Y Series ↓</td><td>Classes</td><td rowspan="2">$Total\ of\ frequencies\ of\ Y$
$g(y)$</td></tr>
<tr><td>Mid Point
$x_1 x_2 \quad \ldots x_i \ldots \quad \ldots . x_m$</td></tr>
<tr><td></td><td>y_1
y_2
.
.
.
y_i
:
:
y_n</td><td>$f(x,y)$</td><td>$g(y) = \sum_x f(x,y)$</td></tr>
<tr><td colspan="2">$Total\ of\ frequencies\ of\ X$
$g(x)$</td><td>$f(x) = \sum_y f(x,y)$</td><td>N
$\rightarrow \sum_x\sum_y f(x,y)$
↓
$\sum_y\sum_x f(x,y)$</td></tr>
</table>

Table: 9.4.2

Problem 9.5.1. The following table gives, according to age, the frequency of marks obtained by 100 students in an intelligence test.

Ages in Years → Marks. ↓	18	19	20	21	**Total**
10-20	4	2	2	-	8
20-30	5	4	6	4	19
30-40	6	8	10	11	35
40-50	4	4	6	8	22
50-60	-	2	4	4	10
60-70	-	2	3	1	6
Total	19	22	31	28	100

Table: 9.4.3

Calculate the correlation coefficient.

Solution. Let $U = X - 19;\ V = \left\{\frac{Y-35}{10}\right\}$

CORRELATION TABLE

v	y(Mid-Value)	Marks u Age (X) Marks (Y)
-2	15	10-20
-1	25	20-30
0	35	30-40
1	45	40-50
2	55	50-60
3	65	60-70

	-1 **18**	**0** **19**	**1** **20**	**2** **21**	**Total** $g(v)$	$vg(v)$	$v^2g(v)$	$\sum_u uv\,f(u,v)$
	4 (8)	2 (0)	4 (-4)		8	-16	32	4
	5 (5)	4 (0)	6 (-6)	4 (-8)	10	-19	19	-9
	6 (0)	8 (0)	10 (0)	11 (0)	35	0	0	0
	4 (-4)	4 (0)	6 (6)	8 (16)	22	22	22	18
		2 (0)	4 (8)	4 (16)	10	20	40	24
		2 (0)	3 (9)	1 (6)	6	18	54	15
Total $f(u)$	19	22	31	28	100	25	167	52
$uf(u)$	-19	0	31	56	68			
$u^2f(u)$	19	0	31	112	162			
$\sum_v uvf(u,v)$	9	0	13	30	52			

Table: 9.4.4

$$\bar{u} = \frac{1}{N}\sum_u uf(u) = \frac{68}{100} = 0.68, \bar{v} = \frac{1}{N}\sum_v vg(v) = \frac{25}{100} = 0.25$$

$$\text{Cov}(u,v) = \frac{1}{N}\sum_u\sum_v uvf(u,v) = \bar{u}\bar{v} \; = \frac{1}{100} \times 52 - 0.68 \times 0.25$$
$$= 0.35$$

$$\sigma_U{}^2 = \frac{1}{N}\sum_u u^2\, f(u) - \bar{u}^2 = \frac{162}{100} - (0.68)^2 = 1.1576$$

$$\sigma_V{}^2 = \frac{1}{N}\sum_v v^2\, g(v) - \bar{v}^2 = \frac{167}{100} - (0.25)^2 = 1.6075$$

$$r(U,V) = \frac{\text{Cov}(U,V)}{\sigma_U \sigma_V} = \frac{0.35}{\sqrt{1.1576 \times 1.6075}} = 0.25$$

Since correlation coefficient is independent of change of origin and scale, $r(X,Y) = r(U,V) = 0.25$

- Figures in circles in the table are the product terms $uvf(u,v)$.

9.6 .RANK CORRELATION:-

Let us suppose that a group of n individuals is arranged in order of merit or proficiency in possession of two characteristics A and B. These ranks in the two characteristics will in general, be different.

For example if we consider the relation between intelligence and beauty; it is not necessary that a beautiful individual is intelligent also. Consider that, for $i = 1,2,3,\ldots,n$ let (x_i, y_i)be the ranks of the ith individual in two characteristics A and B respectively. Karl Pearson's coefficient of correlation between the ranks of X and $Y$is called the rank correlation coefficient between A and B for that group of individuals.

Since (x_i, y_i) be the ranks of the ith individual in two characteristics A and B respectively, therefore each of the variables X and Y takes the values 1.2 $\ldots\ldots n$.

Hence $\bar{x} = \bar{y} = \frac{1}{n}[1 + 2 + 3 + \cdots + n] = \frac{n+1}{2}$ and

$$\sigma_x^2 = \left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right) - \bar{x}^2 = \frac{1}{n}[1^2 + 2^2 + 3^2 + \cdots + n^2] - \left(\frac{n+1}{2}\right)^2$$
$$= \frac{n^2 - 1}{12}.$$

Similarly, $\sigma_y^2 = \frac{n^2-1}{12}$. Define, $d_i = x_i - y_i$. Then $d_i = x_i - \bar{x} - (y_i - \bar{y})$. This implies that $\sum d_i^2 = \sum[(x_i - \bar{x}) - (y_i - \bar{y})]^2$ i.e. $\sum d_i^2 = \sum C + \sum(y_i - \bar{y})^2 - 2\sum(x_i - \bar{x})(y_i - \bar{y})$.This gives $\frac{\sum d_i^2}{n} = \sigma_x^2 + \sigma_y^2 - 2Cov(X,Y) = \sigma_x^2 + \sigma_y^2 - 2\rho\sigma_x\sigma_y$. Here, ρ is the Karl Pearson's coefficient of correlation between the ranks of X_i and Y_i and by definition it is the rank correlation coefficient between A and B. Using $\sigma_x = \sigma_y$, we have

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)}.$$

Problem 9.6.1: The rank of same 16 students in Mathematics and Physics are as follows. Two numbers within brackets denote the ranks of the students in Mathematics and Physics:
(1,1) (2,10) (3,3) (4,4) (5,5) (6,7) (7,2) (8,6) (9,8) (10,11) (11,15) (12,9) (13,14) (14,12) (15,16) (16,13).
Calculate the rank correlation coefficient for proficiencies of this group in Mathematics and Physics.

Solution.

Rank in Math. X	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Total
Rank in Physics Y	1	10	3	4	5	7	2	6	8	11	15	9	14	12	16	13	
$d = X - Y$	0	-8	0	0	0	-1	5	2	1	-1	-4	3	-1	2	-1	3	0
d^2	0	64	0	0	0	1	25	4	1	1	16	9	1	4	1	9	136

Rank correlation coefficient is given by

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)} = 1 - \frac{6 \times 136}{16 \times 255} = 1 - \frac{1}{5} = 0.8$$

Repeated Ranks:

If any two or more individuals are bracketed equal in any classification with respect to characteristic A and B, or if there is more than one item with the same value in the series. In this case, common ranks are given to the repeated items. This common rank is the average of the ranks which these items would have assumed if they were slightly different from each other. The next item will get the rank next to the ranks already assumed.

After that, in the formula of the rank correlation coefficient, we add the factor $\frac{m(m^2-1)}{12}$ to $\sum d^2$, where m is the number of times an item is repeated. This correction factor is to be added for each repeated value in both the X −series. and Y-series.

Problem 9.6.2: Obtain the rank correlation coefficient for the following data:

X	68	64	75	50	64	80	75	40	55	64
Y	62	58	68	45	81	60	68	48	50	70

Solution: In the X-series we see that the value 75 occurs 2 times. The common rank given to these values is 2·5 which is the average of 2 and 3, the ranks which these values would have taken if they were different. The next value 68, then gets the next rank which is 4. Again, we see that value 64 occurs thrice. The common rank given to it is 6 which is the average of 5, 6 and 7. Similarly in the Y-series, the value 68 occurs twice and its common rank is 3·5 which is the average of 3 and 4. As a result of these common rankings, the formula for ρ has to be corrected. We add $\frac{m(m^2-1)}{12}$ to $\sum d^2$ for each value repeated, where m is the number of times a value occurs. In the X -series the correction is to be applied twice, once for the value 75 which occurs twice $(m = 2)$ and then for the value 64 which occurs thrice $(m = 3)$. The total correction for the X-series is

											Total
X	68	64	75	50	64	80	75	40	55	64	
Y	62	58	68	45	81	60	68	48	50	70	
Rank by X	4	6	2.5	9	6	1	2.5	10	8	6	
Rank by Y	5	7	3.5	10	1	6	3.5	9	8	2	
$d = X - Y$	-1	-1	-1	-1	5	-5	-1	1	0	4	$\sum d = 0$
d^2	1	1	1	1	25	25	1	1	0	16	$\sum d^2 = 72$

$$\frac{2(4-1)}{12} + \frac{3(9-1)}{12} = \frac{5}{2}$$

Similarly, this correction for the Y-series is $\frac{2(4-1)}{12} = \frac{1}{2}$, as the value 68 occurs twice. Thus

$$\rho(X,Y) = 1 - \frac{6\left[\sum d^2 + \frac{5}{2} + \frac{1}{2}\right]}{n(n^2-1)} = 1 - \frac{6 \times (72+3)}{10 \times 99} = 0.545$$

9.7 SOLVED EXAMPLES:-

Example 9.7.1. Calculate the correlation coefficient for the following heights (in inches) of father's (X) and their son's (Y):

X	65	66	67	67	68	69	70	72
Y	67	68	65	68	72	72	69	71

Solution: Calculations for correlation coefficient

									Total
X	65	66	67	67	68	69	70	72	544
Y	67	68	65	68	72	72	69	71	552
X^2	4225	4356	4489	4489	4624	4761	4900	5184	37028
Y^2	4489	4624	4225	4624	5184	5184	4761	5041	38132
XY	4355	4488	4355	4556	4896	4968	4830	5112	37560

$$\overline{X} = \frac{1}{n}\sum X = \frac{544}{8} = 68,\ \overline{Y} = \frac{1}{n}\sum Y = \frac{552}{8} = 69$$

$$r(X,\ Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} = \frac{\left(\frac{1}{n}\sum XY - \overline{XY}\right)}{\sqrt{\left(\frac{1}{n}\sum X^2 - \overline{Y}^2\right)\left(\frac{1}{n}\sum Y^2 - \overline{Y}^2\right)}}$$

$$= \frac{\left(\frac{1}{8} \times 37560 - 68 \times 69\right)}{\sqrt{\left(\frac{37028}{8} - 68^2\right)\left(\frac{38132}{8} - 69^2\right)}}$$

$$= \frac{(4695 - 4692)}{\sqrt{(4628.5 - 4624)(4766.5 - 4761)}}$$

$$= 0.603$$

Another Method: Using change of origin. Let $U = X - 68$ &$V = Y - 69$.

									Total
X	65	66	67	67	68	69	70	72	544
Y	67	68	65	68	72	72	69	71	552
U	-3	-2	-1	-1	0	1	2	4	0
V	-2	-1	-4	-1	3	3	0	2	0
U^2	9	4	1	1	0	1	4	16	36
V^2	4	1	16	1	9	9	0	4	44
UV	6	2	4	1	0	3	0	8	24

$$\overline{U} = \frac{1}{n}\sum U = 0,\ \overline{V} = \frac{1}{n}\sum V = 0$$

$$Cov(U,\ V) = \frac{1}{n}\overline{\sum UV} - \overline{U}\,\overline{V} = \frac{1}{8} \times 24 = 3$$

$$\sigma_{U^2} = \frac{1}{n}\sum U^2 - \left(\overline{U}\right)^2 = \frac{1}{8} \times 36 = 4.5$$

$$\sigma_{V^2} = \frac{1}{n}\sum V^2 - \left(\overline{V}\right)^2 = \frac{1}{8} \times 44 = 5.5$$

$$\therefore \quad r(U,\ V) = \frac{Cov(U,\ V)}{\sigma_U \sigma_V} = \frac{3}{\sqrt{4.5 \times 5.5}} = 0.603 = r(X,\ Y)$$

Example 9.7.2.A computer while calculating correlation coefficient between two variables X and Y from 25 pairs of observations obtained the following results:

N=25, $\sum X = 125,\ \sum X^2 = 650,\ \sum Y = 100,\ \sum Y^2 = 460,\ \sum XY = 508$

It was, however, later discovered at the time of checking that he had copied down two pairs as

X	Y
6	14
8	6

While the correct values were

X	Y
8	12
6	8

Obtain the correct value of correlation coefficient.

Solution: The corrected $\sum X, \sum Y, \sum X^2, \sum Y^2, \sum XY, \overline{X}, \overline{Y}$ *are*:

$$\sum X = 125 - 6 - 8 + 8 + 6 = 125$$

$$\sum Y = 100 - 14 - 6 + 12 + 8 = 100$$

$$\sum X^2 = 650 - 6^2 - 8^2 + 8^2 + 6^2 = 650$$

$$\sum Y^2 = 460 - 14^2 - 6^2 + 12^2 + 8^2 = 436$$

$$\sum XY = 508 - 6 \times 14 - 8 \times 6 + 8 \times 12 + 6 \times 8 = 520$$

$$\overline{X} = \frac{1}{n}\sum X = \frac{1}{25} \times 125 = 5,\ \overline{Y} = \frac{1}{n}\sum Y = \frac{1}{25} \times 100 = 4$$

$$Cov\,(X, Y) = \frac{1}{n}\sum XY - \overline{X}\,\overline{Y} = \frac{1}{25} \times 520 - 5 \times 4 = \frac{4}{5}$$

$$\sigma_X^2 = \frac{1}{n}\sum X^2 - \overline{X}^2 = \frac{1}{25} \times 650 - 5^2 = 1$$

$$\sigma_Y^2 = \frac{1}{n}\sum Y^2 - \overline{Y}^2 = \frac{1}{25} \times 436 - 16 = \frac{36}{25}$$

Hence the corrected correlation coefficient is: $r(X, Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} = \frac{\frac{4}{5}}{1 \times \frac{6}{5}} = \frac{2}{3} = 0.67$

Example 9.7.3. Show that if X', Y' are the deviations of the random variables X *and* Y from their respective means then

(i) $r = 1 - \frac{1}{2N}\sum_i \left(\frac{X_i'}{\sigma_X} - \frac{Y_i'}{\sigma_Y}\right)^2$

(ii) $r = -1 + \frac{1}{2N}\sum_i \left(\frac{X_i'}{\sigma_X} + \frac{Y_i'}{\sigma_Y}\right)^2$

Solution. (i) Here $X_i' = \left(x_i - \overline{X}\right)\ and\ Y_i' = \left(Y_i - \overline{Y}\right)$

R.H.S.= $1 - \frac{1}{2N}\sum_i \left(\frac{X_i'}{\sigma_X} - \frac{Y_i'}{\sigma_Y}\right)^2 = 1 - \frac{1}{2N}\left[\frac{1}{\sigma_{X^2}}\sum_i X_i'^2 + \frac{1}{\sigma_{Y^2}}\sum_i Y_i'^2 - \frac{2}{\sigma_X\sigma_Y}\sum_i X_i'Y_i'\right]$

$$= 1 - \frac{1}{2N}\left[\frac{1}{\sigma_{X^2}}\sum_i \left(X_i' - \overline{X}\right)^2 + \frac{1}{\sigma_{Y^2}}\sum_i \left(Y_i - \overline{Y}\right)^2 - \frac{2}{\sigma_X\sigma_Y}\sum_i \left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)\right]$$

$$= 1 - \frac{1}{2}\left[\frac{1}{\sigma_{X^2}} \times \sigma_{X^2} + \frac{1}{\sigma_{Y^2}} \times \sigma_{Y^2} - \frac{2}{\sigma_X\sigma_Y} \times r\sigma_X\sigma_Y\right]$$

$$= 1 - \frac{1}{2}[1 + 1 - 2r] = r$$

(ii) Proceeding similarly, we will get

R.H.S. $= -1 + \frac{1}{2}[1 + 1 + 2r] = r$

Example 9.7.4 The variables X and Y are connected by the equation $aX + bY + c = 0$. Show that the correlation between them is -1 if the signs of a and b are alike and $+1$ if they are different.

Solution. $aX + bY + c = 0 \Rightarrow aE(X) + bE(Y) + c = 0$

$$\therefore \quad a\{X - E(X)\} + b\{Y - E(Y)\} = 0$$

$$\Rightarrow \quad \{X - E(X)\} = \frac{b}{a}\{Y - E(Y)\}$$

$$\therefore \quad Cov(X, Y) = E\left[\left(X - E(X)\right)\left(Y - E(Y)\right)\right]$$

$$= -\frac{b}{a}E\left[\left(Y - E(Y)\right)^2\right]$$

$$= -\frac{b}{a} \times \sigma_Y^2 \qquad E\left[\left(X - E(X)\right)^2\right]$$

$$= \frac{b^2}{a^2}E\left[\left(Y - E(Y)\right)^2\right] = -\frac{b^2}{a^2} \times \sigma_Y^2$$

$$\therefore \quad r = \frac{-\frac{b}{a}\sigma_Y^2}{\sqrt{\sigma_Y^2\sqrt{\frac{b^2}{a^2}\sigma_Y^2}}} = -\frac{\frac{b}{a}\sigma_Y^2}{\left|\frac{b}{a}\right|\sigma_Y^2}$$

Therefore, $r = 1$ if $b\ and\ a$ are of opposite signs and $r = -1,\ if\ b\ and\ a$ are of same sign

Example 9.7.5.$(\boldsymbol{a})$ If $Z = aX + bY\ and\ r$ is the correlation coefficient between $X\ and\ Y,$ show that $\sigma_Z^2 = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2abr\sigma_X\sigma_Y$

$(\boldsymbol{b})$ Show that the correlation coefficient r between two random variables $X\ and\ Y$ is given by

$$r = \frac{(\sigma_X^2 + \sigma_Y^2 - \sigma_{X-Y}^2)}{2\sigma_X\sigma_Y}$$

Where $\sigma_X,\ \sigma_Y\ and\ \sigma_{X-Y}$ are the standard deviations of $X,\ Y\ and\ X - Y\ respectively$.

Solution.(a) Taking expectation of both sides of $Z = aX + bY,$ we get

$$E(Z) = aE(X) + bE(Y),$$

$$\therefore \quad Z - E(Z) = a[X - E(X)] + b[Y - E(Y)]$$

Squaring and taking expectation of both sides, we get

$$\sigma_Z^2 = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2abCov(X,\ Y)$$

$$= a^2\sigma_X^2 + b^2\sigma_Y^2 + 2abr\sigma_X\sigma_Y$$

(b) Taking $a = 1,\ b = -1$ in the above case, we have

$$Z = X - Y\ and\ \sigma_{X-Y}^2 = \sigma_X^2 + \sigma_Y^2 - 2r\sigma_X\sigma_Y$$

Therefore,

$r = \frac{\sigma_X^2 + \sigma_Y^2 - \sigma_{X-Y}^2}{2\sigma_X\sigma_Y}.$

Example 9.7.6. $X\ and\ Y$ are two random variables with variances σ_X^2 and σ_Y^2 respectively and r is the coefficient of correlation between them. If $U = X + kY\ and\ V = X + \frac{\sigma_X}{\sigma_Y}Y,$ find the value of k so that $U\ and\ V$ are uncorrelated .

Solution. Taking expectations of $U = X + kY\ and\ V = X + \frac{\sigma_X}{\sigma_Y}Y$ we get

$$E(U) = E(X) + kE(Y) and E(V) = E(X) + \sigma_X/\sigma_Y E(Y)$$

$$U - E(U) = [X - E(X)] + k[Y - E(Y)]\ and\ V - E(V)$$

$$= [X - E(X)] + \sigma_X/\sigma_Y[Y - E(Y)]$$

$$Cov(U,\ V) = E\big[\big(U - E(U)\big)\big(V - E(V)\big)\big]$$

$$= E\left[\big(X - E(X)\big) + k\big(Y - E(Y)\big) \times \big(X - E(X)\big) + \frac{\sigma_X}{\sigma_Y}\big(Y - E(Y)\big)\right]$$

$$= \sigma_X^2 + \frac{\sigma_X}{\sigma_Y}Cov(X,\ Y) + kCov(\ X,\ Y) + k\frac{\sigma_X}{\sigma_Y} \times \sigma_Y^2$$

$$= [\sigma_X^2 + k\sigma_X\sigma_Y] + \left[\frac{\sigma_X}{\sigma_Y} + k\right]Cov(X,\ Y)$$

$$= \sigma_X(\sigma_X + k\sigma_Y) + \left[\frac{\sigma_X + k\sigma_Y}{\sigma_Y}\right]Cov(X,\ Y)$$

$$= (\sigma_X + k\sigma_Y) + \left[\sigma_X + \frac{Cov(X,\ Y)}{\sigma_Y}\right] = (\sigma_X + k\sigma_Y) + (1 + r)\sigma_X$$

Now, $U\ and\ V$ are uncorrelated if

$$r(U,\ V) = 0\ i.e.Cov(U,\ V) = 0$$

Which means, $(\sigma_X + k\sigma_Y) + (1 + r)\sigma_X = 0$

This implies that

$$\sigma_X + k\sigma_Y = 0$$

And hence,

$k = -\frac{\sigma_X}{\sigma_Y}.$

Example 9.7.7. The random variable $X\ and\ Y$ are jointly normally distributed and $U\ and\ V$ are defined by

$U = X\,cos\alpha + Y\,sin\alpha$,

$$V = Y\,cos\alpha - Y\,sin\alpha$$

Show that $U\ and\ V$ will be uncorrelated if

$$\tan 2\alpha = \frac{2r\sigma_X\sigma_Y}{\sigma_X^2 - \sigma_Y^2},$$

where$r$ is correlation coefficient between X and Y, $\sigma_X^2 = Var(X)\ and\ \sigma_Y^2 = Var(Y)$. Are $U\ and\ V$ independent?

Solution. We have

$$Cov(U,\,V) = E\big[\big(U - E(U)\big)\big(V - E(V)\big)\big]$$

$$= E\Big[\big[\big(X - E(X)\big)cos\alpha + \big(V - E(V)\big)sin\alpha\big]$$

$$\times \big[\big(Y - E(Y)\big)cos\alpha - \quad \big(X - E(X)\big)sin\alpha\big]\Big]$$

$$= \cos^2\alpha\, Cov(X,\,Y) - sin\alpha cos\alpha.\sigma_X^2 + sin\alpha cos\alpha.\sigma_Y^2$$

$$- \sin^2\alpha\,\big(Cov(X,\,Y)\big)$$

$$= (\cos^2\alpha - \sin^2\alpha)Cov(X,\,Y) - sin\alpha cos\alpha(\sigma_X^2 - \sigma_Y^2)$$

$$= \cos 2\alpha.\,Cov(X,\,Y) - sin\alpha cos\alpha(\sigma_X^2 - \sigma_Y^2)$$

Now, $U\ and\ V$ Will be uncorrelated if and only if $r(U,\,V) = 0$ i.e. if and only if $Cov(U,\,V) = 0$. Which further equivalent to

$$\cos 2\alpha\, Cov(X,\,Y) - sin\alpha cos\alpha.(\sigma_X^2 - \sigma_Y^2) = 0$$

$$\cos 2\alpha\, r\sigma_X\sigma_Y = \frac{\sin 2\alpha}{2}.(\sigma_X^2 - \sigma_Y^2)$$

And hence,

$$\tan 2\alpha = \frac{2r\sigma_X\sigma_Y}{\sigma_X^2 - \sigma_Y^2}$$

However, $r(U,\,V) = 0$ does not imply that the variables $U\ and\ V$ are independent.

Example 9.7.8. If $X\ and\ Y$ are standardized random variables, and

$$r(aX + bY,\,bX + aY) = \frac{1 + 2ab}{a^2 + b^2}$$

Find $r(X,\,Y)$, the coefficient of correlation between $X\ and\ Y$.

Solution. Since $X\ and\ Y$ are standard random variables, we have

$$E(X) = E(Y) = 0$$

And $Var(X) = Var(Y) = 1$ implies that $E(X^2) = E(Y^2) = 1$

And $Cov(X,\,Y) = E(XY)$implies that $E(XY) = r(X,\,Y)\sigma_X\sigma_Y = r(X,Y)$

Now,

$$r(aX + bY,\,bX + aY)$$

$$= \frac{E[(aX + bY)(bX + aY)] - E(aX + bY)E(bX + aY)}{[Var(aX + bY)Var(bX + aY)]^{\frac{1}{2}}}$$

$$= \frac{E[abX^2 + a^2XY + b^2YX + abY^2] - 0}{\{[a^2Var(X) + b^2Var(Y) + 2abCov(X,Y)] \times [b^2Var(X) + a^2Var(Y) + 2abCov(X,Y)]\}^{\frac{1}{2}}}$$

$$= \frac{ab + a^2r(X,Y) + b^2r(X,Y) + a}{\{[a^2 + b^2 + 2a \quad (X,Y)][b^2 + a^2 + 2bar(X,Y)]\}^{\frac{1}{2}}}$$

$$= \frac{2ab + (a^2 + b^2)r(X,Y)}{a^2 + b^2 + 2abr(X,Y)}.$$

And hence,

$$\frac{1+2ab}{a^2+b^2}=\frac{2ab+(a^2+b^2)r(X,\,Y)}{a^2+b^2+2ab\,r(X,\,Y)}$$

i.e.

$$(a^2+b^2)(1+2ab)+2ab.r(X,\,Y)(1+2ab)$$
$$=(a^2+b^2)^2.r(X,\,Y)+2ab(a^2+b^2)$$

This implies that

$$(a^4+b^4+2a^2b^2-2ab-4a^2b^2).r(X,\,Y)=(a^2+b^2)$$

Further,

$$[(a^2-b^2)^2-2ab].r(X,\,Y)=a^2+b^2$$

Then,

$$r(X,\,Y)=\frac{a^2+b^2}{(a^2-b^2)^2-2ab}$$

Example 9.7.9. If $U=aX+bY$ and $V=cX+dY$, where $X\ and\ Y$ are measured from their respective means and if r is the correlation coefficient between $X\ and\ Y$, and if $U\ and\ V$ are uncorrelated, show that

$$\sigma_U\sigma_V=(ad-bc)\sigma_X\sigma_Y(1-r^2)^{\frac{1}{2}}$$

Solution. We know that,

$$r=\frac{Cov(X,\,Y)}{\sigma_X\sigma_Y}.$$

Then

$$(1-r^2)=1-\frac{[Cov(X,\,Y)]^2}{\sigma_X^2\sigma_Y^2}$$

Which implies that

$$(1-r^2)\sigma_X^2\sigma_Y^2=\sigma_X^2\sigma_Y^2-[Cov(X,\,Y)]^2 \qquad \dots\dots(*)$$
$$U=aX+bY,\,V=cX+dY$$

Since $X,\,Y$ are measured from their means,
$E(X)=0=E(Y)$. Consequently $E(U)=0=E(V)$.
And then, $\sigma_U^2=E(U^2)$; $\sigma_V^2=E(V^2)$.
Also $aX+bY-U=0\ and\ cX+bY-V=0$. Therefore

$$\frac{X}{-bV+dU}=\frac{Y}{-cU+aV}=\frac{1}{ad-bc}$$

Then value of X and Y are:

$$X=\frac{1}{ad-bc}(dU-bV)$$
$$Y=\frac{1}{ad-bc}(-cU+aV)$$

Also since $U,\,V$ are uncorrelated therefore $Cov(U,\,V)=0$. Thus,

$$Var(X)=\frac{1}{(ad-bc)^2}[d^2\sigma_U^2+b^2\sigma_V^2-2bdCov(U,\,V)]$$
$$=\frac{1}{(ad-bc)^2}[d^2\sigma_U^2+b^2\sigma_V^2]$$

Similarly, we have

$$Var(Y) = \frac{1}{(ad-bc)^2}[c^2\sigma_U^2 + a^2\sigma_V^2]$$

$$Cov(X,\,Y) = E(XY) - E(X)E(Y) = E(XY) \quad \{\because \;\; E(X) = 0 = E(Y)\}$$

$$= \frac{1}{(ad-bc)^2}E[(dU - bV)(-cU + aV)]$$

$$= \frac{1}{(ad-bc)^2}[-cd\sigma_U^2 - ab\sigma_V^2]$$

$$= \frac{-1}{(ad-bc)^2}[cd\sigma_U^2 + ab\sigma_V^2]$$

On substituting in (*), we get

$$(1-r^2)\sigma_X^2\sigma_Y^2$$

$$= \frac{1}{(ad-bc)^4} \times [(d^2\sigma_U^2 + b^2\sigma_V^2)(c^2\sigma_U^2 + a^2\sigma_V^2) - (cd\sigma_U^2 + ab\sigma_V^2)]$$

$$= \frac{1}{(ad-bc)^4} \times [c^2d^2\sigma_U^4 + a^2b^2\sigma_V^4 + (a^2d^2 + b^2c^2)\sigma_U^2\sigma_V^2 - c^2d^2\sigma_U^4 - a^2b^2\sigma_V^4 - 2abcd\sigma_U^2\sigma_V^2]$$

$$= \frac{1}{(ad-bc)^4} \times [-c^2d^2\sigma_U^4 - a^2b^2\sigma_V^4 + (a^2d^2 + b^2c^2)\sigma_U^2\sigma_V^2 + c^2d^2\sigma_U^4 + a^2b^2\sigma_V^4 + 2abcd\sigma_U^2\sigma_V^2]$$

$$= \frac{1}{(ad-bc)^4} \times [(a^2d^2 + b^2c^2 - 2abcd)\sigma_U^2\sigma_V^2]$$

$$= \frac{1}{(ad-bc)^4} \times (ad-bc)^2\sigma_U^2\sigma_V^2$$

And hence,

$$\sigma_U\sigma_V = (ad-bc)\sigma_X\sigma_Y(1-r^2)^{\frac{1}{2}}$$

Example 9.7.10. The independent variables $X \; and \; Y$ are defined by:

$$f(x) = 4ax,\; 0 \le x \le r$$

$= 0,\;\; otherwise$ &

$$f(y) = 4by,\; 0 \le y \le s$$
$$= 0,\;\; otherwise$$

Then show that:

$$Cov(U,\,V) = \frac{b-a}{b+a},$$

where $U = X + Y \quad and \; V = X - Y$

Solution . Since the total area under probability curve is unity (one), we have:

$\int_0^r f(x)dx = 4a\int_0^r xdx = 1$ implies that $2ar^2 = 1$

Which gives,

$$a = \frac{1}{2r^2} \qquad \text{... (9.7.1)}$$

Similarly, $\int_0^r f(y)dy = 4b\int_0^r ydy = 1$ gives, $b = \frac{1}{2s^2}$ ... $(\mathbf{9.7.2})$

Therefore,

$$f(x) = 4ax = \frac{2x}{r^2},\ \ 0 \le x \le r\,;$$

And

$$f(y) = 4by = \frac{2y}{s^2},\ \ 0 \le y \le s \qquad ...(\mathbf{9.7.3})$$

Since $X\ and\ Y$ are independent variates, therefore

$r(X,\,Y) = 0$ which further imlies that $Cov(X,\,Y) = 0$ ... $(\mathbf{9.7.4})$

Now,

$$Cov(U,V) = Cov(X+Y,\,X-Y)$$

$$= Cov(X,\,X) - Cov(X,\,Y) + Cov(Y,\,X) - Cov(Y,\,Y)$$

$= \sigma_X^2 - \sigma_Y^2$ [Using equation (9.7.4)]

And

$$Var(U) = Var(X) + Var(Y) + 2Cov(X,\,Y)$$

$= \sigma_X^2 + \sigma_Y^2$[Using equation (9.7.4)]

$$Var(Y) = Var(X-Y) = Var(X) + Var(Y) - 2Cov(X,\,Y)$$

$= \sigma_X^2 + \sigma_Y^2$[Using equation (9.7.4)]

$$\therefore \qquad r(U,\,V) = \frac{Cov(U,\,V)}{\sigma_U\sigma_V} = \frac{\sigma_X^2 - \sigma_Y^2}{\sigma_X^2 + \sigma_Y^2}$$

Further,

$$E(X) = \int_0^r xf(x)dx = \frac{2}{r^2}\int_0^r x^2\,dx = \frac{2r}{3}$$

$$E(X^2) = \int_0^r x^2\,f(x)dx = \frac{2}{r^2}\int_0^r x^3\,dx = \frac{r^2}{2}$$

$$Var(X) = E(X^2) - [E(X)]^2 = \frac{r^2}{2} - \frac{4r^2}{9} = \frac{r^2}{18} = \frac{1}{36a}$$

Similarly,

$E(Y) = \frac{2s}{3}$, $E[Y^2] = \frac{s^2}{2}$ and $Var(Y) = \frac{s^2}{18} = \frac{1}{36b}$

On substituting in (v) we get

$$r(U,\,V) = \frac{\frac{1}{(36a)} - \frac{1}{(36b)}}{\frac{1}{(36a)} + \frac{1}{(36b)}} = \frac{b-a}{b+a}$$

Example9.7.11. Let the random variable X have the marginal density

$f_1(x) = 1,\ -\frac{1}{2} < x < \frac{1}{2}$and let the conditional density of Y be

$$f(y|x|) = 1,\ x < y < x+1,\ -\frac{1}{2} < x < 0$$

$$= 1,\ -x < y < -x+1,\ 0 < x < \frac{1}{2}$$

Show that the variables $X\ and\ Y$ are uncorrelated.

Solution. We have

$$E(X) = \int_{-\frac{1}{2}}^{\frac{1}{2}} xf_1(x)dx = \int_{-\frac{1}{2}}^{\frac{1}{2}} x\,.\,1dx = \left|\frac{x^2}{2}\right|_{-\frac{1}{2}}^{\frac{1}{2}} = 0$$

If $f(x,\ y)$ *is the joint p.d.f of X and Y*, then

$$f(x,y) = f(y|x)f_1(x) = f(y|x) \qquad [\because\ f_1(x) = 1]$$

$$E(XY) = \int_{-\frac{1}{2}}^{0}\int_{x}^{x+1} xyf(x,y)dxdy + \int_{0}^{\frac{1}{2}}\int_{-x}^{1-x} xyf(x,y)dxdy$$

$$= \int_{-\frac{1}{2}}^{0}\left[x\int_{x}^{x+1} ydy\right]dx + \int_{0}^{\frac{1}{2}}\left[x\int_{-x}^{1-x} ydy\right]dx$$

$$= \frac{1}{2}\left[\frac{2}{3}x^3 + \frac{x^2}{2}\right]_{-\frac{1}{2}}^{0} + \frac{1}{2}\left[\frac{x^2}{2} - \frac{2}{3}x^3\right]_{0}^{\frac{1}{2}}$$

$$= \frac{1}{2}\left[\frac{1}{12} - \frac{1}{8} - \frac{1}{12} + \frac{1}{8}\right] = 0.$$

i.e. X & Y are uncorrelated.

Example 9.7.12: Ten competitors in a musical test were ranked by the three judges A. B and C in the following order:

Rank by A	1	6	5	10	3	2	4	9	7	8
Rank by B	3	5	8	4	7	10	2	1	6	9
Rank by C	6	4	9	8	1	2	3	10	5	7

Using rank correlation method, discuss which pair of judges has the nearest approach to common liking in music.

Solution: Here $n = 10$

											Total
Rank by A (X)	1	6	5	10	3	2	4	9	7	8	
Rank by B (Y)	3	5	8	4	7	10	2	1	6	9	
Rank by C (Z)	6	4	9	8	1	2	3	10	5	7	

$d_1 = X - Y$	-2	1	-3	6	-4	-8	2	8	1	-1	
$d_2 = X - Z$	-5	2	-4	-2	2	0	1	-1	2	1	
$d_3 = Y - Z$	-3	1	-1	-4	6	8	-1	-9	1	2	
d_1^2	4	1	9	36	16	64	4	64	1	1	$\sum d_1^2 = 200$
d_2^2	25	4	16	4	4	0	1	1	4	1	$\sum d_2^2 = 60$
d_3^2	9	1	1	16	36	64	1	81	1	4	$\sum d_3^2 = 214$

Then the rank correlation coefficients are

$$\rho(X,Y) = 1 - \frac{6\sum d_1^2}{n(n^2-1)} = 1 - \frac{6 \times 200}{10 \times 99} = 1 - \frac{40}{33} = -\frac{7}{33}$$

$$\rho(X,Z) = 1 - \frac{6\sum d_2^2}{n(n^2-1)} = 1 - \frac{6 \times 60}{10 \times 99} = 1 - \frac{4}{11} = \frac{7}{11}$$

$$\rho(Y,Z) = 1 - \frac{6\sum d_3^2}{n(n^2-1)} = 1 - \frac{6 \times 214}{10 \times 99} = -\frac{49}{165}$$

Since $\rho(X,Z)$ is maximum, we conclude that the pair of judges A and C has the nearest approach to common likings in music.

9.8.SUMMARY: -

In this unit we have studied that: how we can relate the variables in bivariate data. We have also studied the correlation coefficient for frequency bivariate distribution. Further we have studied rank correlation, even in the case when ranks are repeating.

CHECK YOUR PROGRESS

Problem1:. Correlation coefficient lies between -1 and +1. True\False

Probelm2: If ρ (x, y) = +1, then there is a perfect negative correlation between x and y. True\False

Probelm3: The correlation coefficient is independent of the change of origin and scale. True\False

Probelm4: The correlation is perfect positive if r = ……….

Probelm5: The correlation coefficient between x and a – x is…………

Probelm6: If the amount of change in one variable tends to be a constant ratio to the amount of change in the other variable than the correlation is said to be……….

Probelm7: The coefficient of correlation from the following points of observation (1, 3), (2, 2), (3, 5), (4, 4), (5, 6) is

a) 1.8 b) 0.8 c) 0 d) -0.8

Probelm8: The coefficient of correlation between x and y for the following data is

X	65	66	67	67	68	69	70	72
Y	67	68	65	68	72	72	69	71

a) 0.604 b) 0.8 c) 1.604 d) 1

Probelm9: From the data given below, the number of items n.

$r = 0.5$, $\sum xy = 120$, $\sum x^2 = 90$, $\sigma_y = 8$, where x and y are deviations from arithmetic mean.

a) 5 b) 8 c) 10 d) 20

Probelm10: There is no correlation between two variables x and y if the value of ρ(x, y) =

a) 1 b) -1 c) 2 d) 0

9.9. GLOSSARY:-

(i). Bivariate data
(ii). Scatter diagram
(iii)Correlation Coefficient
(iv)Rank Coefficient

9.10.REFERENCE BOOKS:-

1. S. C. Gupta and V. K. Kapoor, (2020), *Fundamentals of mathematical statistics*, Sultan Chand & Sons.
2. Seymour Lipschutz and John J. Schiller, (2017), *Schaum's Outline: Introduction to Probability and Statistics*, McGraw Hill Professional.

3. J. S. Milton and J. C. Arnold , (2003), *Introduction to Probability and Statistics (4th Edition),* Tata McGraw-Hill.
4. https://www.wikipedia.org.

9.11.SUGGESTED READINGS:-

1. A.M. Goon,(1998), *Fundamental of Statistics 7th Edition),* 1998.
2. R.V. Hogg and A.T. Craig, (2002), *Introduction to Mathematical Statistics,* MacMacMillan, 2002.
3. Jim Pitman, (1993), *Probability,* Springer-Verlag.
4. https://archive.nptel.ac.in/courses/111/105/111105090

9.12.TERMINAL QUESTIONS:-

TQ 9.12.1 The following are the marks obtained by 24 students in a class test of Statistics and Mathematics:

Role No. of Students	01	02	03	04	05	06	07	08	09	10	11	12
Marks in Statistics	15	00	01	03	16	02	18	05	04	17	06	19
Marks in Mathematics	13	01	02	07	08	09	12	09	17	16	06	18
Roll No. of Students	13	14	15	16	17	18	19	20	21	22	23	24
Marks in Statistics	14	09	08	13	10	13	11	11	12	18	09	07
Marks in Mathematics	11	03	05	04	10	11	14	07	18	15	15	03

Prepare a correlation table taking the magnitude of each class interval as four marks and the first class interval as "equal to 0 and less than 4". Calculate Karl's Pearson's coefficient of correlation between the marks in Statistics and marks in Mathematics from the correlation table.

TQ 9.12.3. The joint probability distribution of X and Y is given below:

	X:	-1	+1
Y:			
0		$\frac{1}{8}$	$\frac{3}{8}$
1		$\frac{2}{8}$	$\frac{2}{8}$

Find the correlation coefficient between X and Y.

9.13 ANSWER:-

Answer of Check your Questions:-

CHQ.1 True
CHQ.2 False
CHQ.3 True
CHQ.4 +1
CHQ.5 -1
CHQ.6 Linear
CHQ.7 b
CHQ.8 a
CHQ.9 c
CHQ.10 d

Answer of Terminal Questions:-
TQ9.12.1 0.5544
TQ9.12.3 -0.2582

UNIT 10:-REGRESSION

CONTENTS:

10.1. INTRODUCTION:-

The term "regression" literally means "stepping back towards the average". It was first used by a British biometrician Sir Francis Galton $(\mathbf{1822-1911})$ in connection with the inheritance of stature. Galton found that the offspring of abnormally tall or short parents tend to "regress" or "step back" to the average population height. But the term "regression" as now used in Statistics is only a convenient term without having any reference to biometry. Regression gives an idea of relationship between two or more variables.

10.2.OBJECTIVES:-

After studying this unit learner will be able to:

1. Analyse the connection between two or more variables.
2. Compute the line of regression.
3. Compute the regression coefficients.

10.3.LINE OF REGRESSION, REGRESSION COEFFICIENTS AND PROPERTIES:-

Definition. Regression analysis is a mathematical measure of the average relationship between two or more variables in terms of the original units of the data. In regression analysis there are two types of variables. The variable whose value is influenced or is to be predicted is called dependent variable and the variable which influences the values or is used for prediction, is called independent variable. In regression analysis independent variable is also known as regressor or predictor or explanatory variable while the dependent variable is also known as regressed or explained variable.

Lines of Regression. If the variable in a bivariate distribution are related, we will find that the points in the scatter diagram will cluster round some curve called the "curve of regression". If the curve is a straight line, it is called the line of regression and there is said to be linear regression between the variables, otherwise regression is said to be curvilinear. The line of regression is the line which gives the best estimate to the value of one variable for any specific value of the other variable. Thus, the line of regression is the line of "best fit" and is obtained by the principles of least squares.

Let us suppose that in the bivariate distribution $(x_i, y_i); i = 1,2, \dots, n$; Y is dependent variable and X is dependent variable. Let the line of regression of Y on X be $Y = a + bX$.
According to the principle of least squares, the normal equations for estimating a and b are

$$\sum_{i=1}^{n} y_i = na + b\sum_{i=1}^{n} x_i \qquad \dots.(\mathbf{10.3.1})$$

$$\sum_{i=1}^{n} x_i\, y_i = a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} {x_i}^2 \qquad .(\mathbf{10.3.2})$$

After dividing by n in the first equation, we get

$$\bar{y} = a + b\bar{x} \qquad \dots.(\mathbf{10.3.2a})$$

Thus, the line of regression of Y and X passes through the point $(\bar{x}, \bar{y})$.
Now, $\mu_{11} = \mathrm{Cov}(X, Y) = \frac{1}{n}\sum_{i=1}^{n} x_i\, y_i - \bar{x}\bar{y}$. This gives

$$\frac{1}{n}\sum_{i=1}^{n} x_i\, y_i = \mu_{11} + \bar{x}\bar{y} \qquad \dots.(\mathbf{10.3.3})$$

Also, ${\sigma_x}^2 = \frac{1}{n}\sum_{i=1}^{n} {x_i}^2 - \bar{x}^2$
Therefore,

$$\frac{1}{n}\sum_{i=1}^{n} x_i^2 = \sigma_x^2 + \bar{x}^2 \qquad \ldots.(\mathbf{10.3.4})$$

Dividing **(10.3.1)** by and using **(10.3.3)** and **(10.3.4),** we get

$$\mu_{11} + \bar{x}\bar{y} = a\bar{x} + b(\sigma x^2 + \bar{x}^2) \qquad \ldots\ldots\ldots\ldots\ldots..(\mathbf{10.3.5})$$

Multiplying ***(10.3.2a)*** by $\bar{x}$ and then subtracting from ***(10.3.5),*** we get

$$\mu_{11} = b\sigma_x^2$$

And hence, $b = \frac{\mu_{11}}{\sigma_x^2}$. Since 'b' is the slope of the line of regression of Y on X and since the line of regression passes through the point $(\bar{x}, \bar{y})$, its equation is

$$Y - \bar{y} = b(X - \bar{x}) = \frac{\mu_{11}}{\sigma_x^2}(X - \bar{x}) \qquad \ldots\ldots\ldots..(\mathbf{10.3.6})$$

i.e. $$Y - \bar{y} = r\frac{\sigma_y}{\sigma_x}(X - \bar{x}) \qquad \ldots\ldots\ldots\ldots...(\mathbf{10.3.6a})$$

Starting with the equation $X = A + BY$ and proceeding similarly or by simply interchanging the variables X and Y in ***(10.3.6)*** and ***(10.3.6a),*** the equation of the line of regression of X on Y becomes $X - \bar{x} = \frac{\mu_{11}}{\sigma_x^2}(Y - \bar{y}), X - \bar{x} = r\frac{\sigma_x}{\sigma_y}(Y - \bar{y})$

Question 10.3.1: Using the least square method obtain the equation of line of regression of Y on X.

Solution: The straight line is defined by

$$Y = a + bX$$

And satisfying the residual (least square) condition

$$S = E[(Y - a - bX)^2] = \text{Minimum}$$

For variations in aand b, is called th line of regression of Y and X.

The necessary and sufficient conditions for a minima of S, subject to variations in a and b are:

(i) $\frac{\partial S}{\partial a} = 0, \frac{\partial S}{\partial b} = 0$ and

(ii) $\Delta = \begin{vmatrix} \frac{\partial^2 S}{\partial a^2} & \frac{\partial^2 S}{\partial a\, \partial b} \\ \frac{\partial^2 S}{\partial b\ \partial a} & \frac{\partial^2 S}{\partial b^2} \end{vmatrix} > 0$ and $\frac{\partial^2 S}{\partial a^2} > 0$

Using condition (i), we get

$$\frac{\partial S}{\partial a} = -2E[\,Y - a - bX] = 0$$

$$\frac{\partial S}{\partial b} = -2E[\,Y - a - bX] = 0$$

This gives, $E(Y) = a + bE(X)$and $E(XY) = aE(X) + bE(X^2)$. The first equation implies that the line of regression of Yon X, passes through the point $(E(X), E(Y))$.

Multiplying $E(Y) = a + bE(X)$by $E(X)$ and subtracting from $E(XY) = aE(X) + bE(X^2)$, we get

$$E(XY) - E(X)E(Y) = b[E(X^2) - [E(X)]^2].$$

This gives, $\text{Cov}(X, Y) = b.\sigma_x^2$ and consequently,

$$b = \frac{Cov(X,\ Y)}{\sigma_x{}^2} = \frac{r\sigma_y}{\sigma_x}$$

Using above equation and subtracting $E(Y) = a + bE(X)$from $Y = a + bX$, we get required equation of line of regression:

$$Y - E(Y) = \frac{Cov(X,\ Y)}{\sigma_x{}^2}\big(X - E(X)\big)$$

i.e. $Y - E(Y) = \frac{r\sigma_y}{\sigma_x}\big(X - E(X)\big)$

CYP 10.3.1: Using the least square method obtain the equation of line of regression of X on Y.

(Hint: The straight line defined by $E = A + BY$and satisfying the residual condition

$$E[X - A - BY]^2 = \text{Minimum},$$

for the variable A and B, is called the line of regression of X on Y.)

Remarks 10.3.1.(i) We note that$\frac{\partial^2 S}{\partial a^2} = 2$ (+ve),$\frac{\partial^2 S}{\partial b^2} = 2E(X^2)$and $\frac{\partial^2 S}{\partial a\,\partial b} = 2E(X)$. Which give

$$\Delta = \frac{\partial^2 S}{\partial a^2}\frac{\partial^2 S}{\partial b^2} - \left(\frac{\partial^2 S}{\partial a\,\partial b}\right)^2$$

$$= 4[E(X^2) - (E(X))^2] = 4\,\sigma_x{}^2 > 0$$

Hence the solution of the least square equations provides a minima of S .

(ii) The regression equation implies that the line of regression of Y on X passes through the point $(\bar{x}, \bar{y})$. Similarly that the line of regression of X on Y also passes through the point $(\bar{x}, \bar{y})$. Hence both the lines of regression pass through the point $(\bar{x}, \bar{y})$. In other words, the mean values $(\bar{x}, \bar{y})$ can be obtained as the point of intersection of the two regression lines.

(iii) There are always two lines of regression, one of Y on X and the other of Xon Y. The line of regression of Y on X is used to estimate or predict the value of Y for any given value of X, i.e., when Y is a dependent variable and X is an independent variable. The estimate so obtained will be best in the sense that it will have the minimum possible error as defined by the principle of least squares. We can also obtain an estimate of X for any given value of Y by using regression line Y on X, but the estimated value obtained will not be best since regression line Y on X is obtained on minimising the sum of the squares of errors of estimates in Y and not in X. Hence to estimate or predict X for any given value of Y, we use the regression line of X on Y, which is derived on minimising the sum of the squares of errors of estimates in X. Here X is a dependent variable and Y is an independent variable. The two regression equations are not reversible or interchangeable because of the simple reason that the basis and assumptions for deriving these equations are quite different. The regression equation of Y on X is obtained on minimizing the sum of the squares of the errors parallel to the Y -axis while the regression equation of X on Y is obtained on minimising the sum of squares of the errors parallel to the X -axis. In a particular case of perfect correlation, positive or negative, i.e., r=± 1, the equation of line of regression of Y on X becomes:

$$Y - \bar{y} = \pm \frac{\sigma_y}{\sigma_x}(X - \bar{x})$$

This implies that

$$\frac{Y - \bar{y}}{\sigma_y} = \pm \left(\frac{X - \bar{x}}{\sigma_x}\right)$$

Similarly, the equation of the line of regression of X on Y becomes :

$$X - \bar{x} = \pm \frac{\sigma_y}{\sigma_x}(Y - \bar{y})$$

i.e.

$$\frac{Y - \bar{y}}{\sigma_y} = \pm \left(\frac{X - \bar{x}}{\sigma_x}\right)$$

Which are same in both situations. Hence in case of perfect correlation, (r = ± 1), both the lines of regression coincide. Therefore, in general, we always have two lines of regression except in the particular case of perfect correlation when both the lines coincide and we get only one line.

10.4.ANGLE BETWEEN TWO LINES OF REGRESSION:-

Equations of the lines of regression of Y on X, and X on Y are

$$Y - \bar{y} = r.\frac{\sigma_y}{\sigma_x}(X - \bar{x}) \text{and} X - \bar{x} = r.\frac{\sigma_x}{\sigma_y}(Y - \bar{y})$$

Slopes of these lines are $r.\frac{\sigma_y}{\sigma_x}$and $\frac{\sigma_y}{r.\sigma_x}$ respectively, If Θ is the angle between the two lines of regression then

$$tan\theta = \frac{r.\frac{\sigma_y}{\sigma_x} \sim \frac{\sigma_y}{r.\sigma_x}}{1 + r.\frac{\sigma_y}{\sigma_x} . \frac{\sigma_y}{r.\sigma_x}} = \frac{r^2 \sim 1}{r}\left(\frac{\sigma_x \sigma_y}{{\sigma_x}^2 + {\sigma_y}^2}\right)$$

$$= \frac{1 - r^2}{r}\left(\frac{\sigma_x \sigma_y}{{\sigma_x}^2 + {\sigma_y}^2}\right) (\because r^2 \leq 1)$$

$$\therefore \ \theta = \tan^{-1}\left\{\frac{1 - r^2}{r}\left(\frac{\sigma_x \sigma_y}{{\sigma_x}^2 + {\sigma_y}^2}\right)\right\}$$

Case (i) if $r = 0, \tan\theta = \infty \Longrightarrow \theta = \frac{\pi}{2}$. Thus if the two variables are uncorrelated, the lines of regression become perpendicular to each other.

Case (ii) if $r = \pm 1$, $\tan\theta = \infty \Longrightarrow \theta = 0$ or π.

In this case the two lines of regression either coincide or they are parallel to each other. But since both the lines of regression pass through the point $(\bar{x}, \bar{y})$, they cannot be parallel. Hence in the case of perfect correlation, positive or negative, the two lines of regression coincide.

Remarks 10.4.1.(i) Whenever two lines intersect, there are two angles between them, one acute angle and the other obtuse angle. Further $\tan\theta > 0$ if $0 < \theta < \frac{\pi}{2}$, $i.e., \theta$ is an acute angle and $\tan\theta < 0$ if $\frac{\pi}{2} < \theta < \pi$, $i.e., \theta$ is an obtuse angle and since $0 < r^2 < 1$, the acute angle (θ_1) and abuse angle θ_2 between the two lines of regression are given by

$$\theta_1 = \text{Acute angle} = \tan^{-1}\left\{\frac{\sigma_x\sigma_y}{{\sigma_x}^2 + {\sigma_y}^2} \cdot \frac{1-r^2}{r}\right\}, \qquad r > 0$$

and $$\theta_2 = \text{Obtuse angle} = \tan^{-1}\left\{\frac{\sigma_x\sigma_y}{{\sigma_x}^2 + {\sigma_y}^2} \cdot \frac{r^2-1}{r}\right\}, \qquad r > 0$$

(ii) When $r = 0, i.e.,$ variables X and Y are uncorrelated, then the lines of regressions of Y on X and X on $Y$are given respectivel by $Y = \bar{Y}$and $X = \bar{X}$.

(iii) The fact that if $r = 0$ (variables uncorrelated), the two lines of regression are perpendicular to each. And if $r = \pm 1$, $\theta = 0$, $i.e.,$ the two lines coincide, leads us to the conclusion that for higher degree of correlation between the variables the angle between the lines is smaller, $i.e.,$ the two lines of regression are nearer to each other. On the other hand, if the lines of regression make a larger angle, they indicate a poor degree of correlation between the variables and ultimately for $\theta = \frac{\pi}{2}$, $r = 0$ $i.e.,$ the lines become perpendicular if no correlation exists between the variable. Thus, by plotting the lines of regression on a graph paper, we can have an approximate idea about the degree of correlation between the two variables under study.

10.5. STANDARD ERROR OF ESTIMATE OR RESIDUAL VARIANCE:-

The equation of the line of regression of Y on X is

$$Y = \bar{Y} + r\frac{\sigma_y}{\sigma_x}(X - \bar{X})$$

In other words, $\frac{Y-\bar{Y}}{\sigma_y} = r\left(\frac{X-\bar{X}}{\sigma_x}\right)$

The residual variance ${S_y}^2$ is the expected value of the squares of deviations of te observed values of Y from the expected values as given by the line of regression of Y on X. Thus,

$${S_y}^2 = E\left[\left\{\bar{Y} + r\,\sigma_y\left(X - \frac{\bar{X}}{\sigma_x}\right)\right\}\right]^2$$

$$= {\sigma_y}^2 E\left\{\frac{Y-\bar{Y}}{\sigma_y} = r\left(\frac{X-\bar{X}}{\sigma x}\right)\right\}^2 = {\sigma_y}^2 E(Y^* - rX^*)^2$$

Where $Y^*$and X^* are standardized variates so that
E$(X^{*2}) = 1 =$ E(Y^{*2}) and E$(X^*Y^*) = r$.

$$\therefore \quad {S_y}^2 = {\sigma_y}^2[E(Y^{*2}) + r^2E(X^{*2}) - 2rE(X^*Y^*)]$$
$$= {\sigma_y}^2(1 - r^2)$$

$$S_y = \sigma_y(1-r^2)^{\frac{1}{2}}$$

Similarly, the standard error of estimate of X is given by

$$S_x = \sigma_x(1-r^2)^{\frac{1}{2}}$$

Remarks 10.5.1(i) Since${S_x}^2$ or ${S_y}^2 \geq 0$ it follows that

$$(1 - r^2) \geq 0 \Rightarrow |r| \leq 1$$

Hence $$-1 \leq r(X, Y) \leq 1$$

(ii) If r=± 1, $S_x = S_y = 0$ so that each deviation is zero, and the two lines of regression ore coincident.

(iii) Since, as $r^2 \to 1, S_x$ and $S_y \to 0$, it follows that departure of the value r^2 from unity indicates the departure of the relationship between the variables X and Y from linearity.

(iv) From the definition of linear regression, the minima condition implies that$S_x$orS_y is the minimum variance.

10.6.CURVILINEAR REGRESSION & REGRESSION CURVES:-

Regression Curves. In modern terminology, the conditional mean $E(Y \mid X = x)$ for a continuous distribution is called the regression function of Y on X and the graph of this function of x is known as the regression curve of Y on X or sometimes the regression curve for the mean of Y. Geometrically, the regression function represents the y co-ordinate of the centre of mass of the bivariate probability mass in the infinitesimal vertical strip bounded by x and $x + dx$.

Similarly, the regression function of X on Y is $E(X| Y = y)$ and the graph of this function of y is called the regression curve (of the mean) of X on Y.

In case a regression curve is a straight line, the corresponding regression is said to be *linear*. If one of the regressions is linear, it does not however follow that the other is also linear.

Theorem10.6.1 . Let(X, Y) be a two-dimensional random variable with $E(X) = \bar{X}$, $E(Y) = \bar{Y}, V(X) = {\sigma_X}^2$, $V(Y) = {\sigma_Y}^2$and let$r = r(X, Y)$ be the correlation coefficient between X and Y. If the regession of Y on X is linear then

$$E(Y \mid X) = \bar{Y} + r\frac{\sigma_y}{\sigma_x}(X - \bar{X})$$

Similarly, If the regression of X on Y is linear, then

$$E(X|Y) = \bar{X} + r\frac{\sigma_x}{\sigma_y}(Y - \bar{Y})$$

Proof. Let the regression equation Y on X be

$$E(Y|x) = a + bx \quad ....(\mathbf{10.6.1})$$

But by definition,

$$E(Y|x) = \int_{-\infty}^{\infty} yf(y|x)dy = \int_{-\infty}^{\infty} y\frac{f(x,y)}{f_x(x)}dy$$

$\therefore$ $$\frac{1}{f_x(x)}\int_{-\infty}^{\infty} yf(x,y)dy = a + bx$$

Multiplying both sides of (2) by $f_x(x)$ and integrating w.r.t. x, we get

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(x,y)dydx = a\int_{-\infty}^{\infty} f_x(x)\,dx + b\int_{-\infty}^{\infty} xf_x(x)dx$$

This gives,

$$\int_{-\infty}^{\infty} y\ \ [\int_{-\infty}^{\infty} f(x,y)dx]dy = a + bE(X)$$

This implies that,

$$\int_{-\infty}^{\infty} yf_y(y)\,dy = a + bE(X)$$

$i.e.,$ $$E(Y) = a + bE(X))\text{ or }\bar{Y} = a + b\bar{\boldsymbol{X}}\,..(\mathbf{10.6.2})$$

Multiplying both sides of (2) by $xf_x(x)$ and integrating w.r.t. x, we get

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy(x,y)dydx = a\int_{-\infty}^{\infty} xf_x(x)\,dx + b\int_{-\infty}^{\infty} x^2 f_x(x)dx$$

This implies that,

$$E(XY) = aE(X) + bE(X^2)$$

Since, $\mu_{11} = E(XY) - E\,(X)E(Y)\ = E(XY) - \ \bar{X}\,\bar{Y}$and $\sigma_x{}^2 = E(X^2) - \{E(X)\}^2 = E(X^2) - \bar{X}^2$, therefore above equation becomes:

$$\mu_{11} + \bar{X}\,\bar{Y} = a\bar{X} + b(\sigma_x{}^2 + \bar{X}^2) \quad ....(\mathbf{10.6.3})$$

Solving ***(10.6.2***) and ***(10.6.3)*** simultaneously, we get

$$b = \frac{\mu_{11}}{\sigma_x{}^2}\ \ and\ a = \ \bar{Y} - \frac{\mu_{11}}{\sigma_x{}^2}\bar{X}$$

Substituting in ***(10.6.1)*** and simplifying, we get the required equation of the line of regression Y on X as

$$E(Y\mid x\,) = \ \bar{Y} + \frac{\mu_{11}}{\sigma_x{}^2}(x - \bar{X})$$

Hence,

$E(Y\mid X) = \ \bar{Y} + \frac{\mu_{11}}{\sigma_x{}^2}(X - \bar{X})$ or $E(Y\mid X) = \ \bar{Y} + r\frac{\sigma_y}{\sigma_x}(X - \bar{X})$.

Similarly, by starting with the line $E(X\mid y) = A + By$ and proceeding similarly we shall obtain the equation of the line of regression of X on Y as $E(X\mid Y) = \bar{X} + \frac{\mu_{11}}{\sigma_x{}^2}(Y - \bar{Y}) = \bar{X} + r\frac{\sigma_x}{\sigma_y}(Y - \bar{Y})$.

10.7. REGRESSION COEFFICIENTS:-

The slope of the line of regression of Y on X is also called the coefficient of regression of Yon X. It represents the increment in the value of dependent variable Y corresponding to a unit change in the value of independent variable . More precisely, we write Regression coefficient of Y on X as b_{yx}. i.e.

$$b_{yx} = \frac{\mu_{11}}{{\sigma_x}^2} = r\frac{\sigma_x}{\sigma_y} \quad \text{....}\,(\mathbf{10.7.1})$$

Similarly, the coefficient of regression of X on Y indicates the change in the value of variable X corresponding to a unit change in the value of variable Y and is given by

$$b_{xy} = \frac{\mu_{11}}{{\sigma_y}^2} = r\frac{\sigma_y}{\sigma_x} \quad \text{....}\,(\mathbf{10.7.2})$$

Problem (10.7.1): Correlation coefficient is the geometric mean between the regression coefficients.

Solution: Multiplying equations ***(10.7.1)*** and ***(10.7.2),*** we get

$$b_{xy} \times b_{yx} = r\frac{\sigma_x}{\sigma_y} \times r\frac{\sigma_y}{\sigma_x} = r^2$$

$$\therefore \qquad r = \pm\sqrt{b_{xy} \times b_{yx}}$$

Remark (10.7.1). We have$r = \frac{\mu_{11}}{\sigma_x . \sigma_x}$, $b_{yx} = \frac{\mu_{11}}{{\sigma_x}^2}$and $b_{xy} = \frac{\mu_{11}}{{\sigma_y}^2}$. It may be noted that the sign of correlation coefficient is the same as that of regression coefficients, since the sign of each depends upon the co-variance term μ_{11}. Thus, if the regression coefficients ae positive, ‘r’ is positive and if the regression coefficients are negative ‘r’ is negative.

Problem (10.7.2): If one of the regression coefficients is greater than unity, the other must be less than unity.

Solution: Let one of the regression coefficients (say) b_{yx} be greater than unity, then we have to show that $b_{xy} < 1$. Now,$b_{yx} > 1$ Hence,$\frac{1}{b_{yx}} < 1$ Also, we know that$r^2 \leq 1$ which means $b_{xy}\ .\ b_{yx} \leq 1$Thus,$b_{xy} \leq \frac{1}{b_{yx}} < 1$.

Problem (10.7.3): Arithmetic mean of the regression coefficients is greater than the correlation coefficient r, provided r> 0.

Solution: We have to prove that $\frac{1}{2}(b_{yx} + b_{xy}) \geq r$

or $\frac{1}{2}\left(r\frac{\sigma_y}{\sigma_x} + r\frac{\sigma_x}{\sigma_y}\right) \geq r$ or $\frac{\sigma_y}{\sigma_x} + \frac{\sigma_x}{\sigma_y} \geq 2$

$(\therefore r > 0)$

This implies that

$${\sigma_y}^2 + {\sigma_x}^2 - 2\,\sigma_y\sigma_x \geq 0 \qquad i.e., \ (\sigma_y - \sigma_x)^2 \geq 0$$

Which is always true, the square of a real quantity is ≥ 0.

Problem (10.7.4): Show that r egression coefficients are independent of the change of origin but not of scale.

Solution: Let $U = \frac{X-a}{h}, \ V = \frac{Y-b}{k} \Longrightarrow \quad X = a + hU, \ Y = b + kV.$
Where $a, b, h, = (> 0)$and $k(> 0)$ are constant.

Then Cov $(X, Y) = hk$ Cov $(U, \ V), \sigma_x{}^2 = h^2 \sigma_u{}^2$ and $\sigma_y{}^2 = k^2 \sigma_v{}^2$

$$b_{yx} = \frac{\mu_{11}}{\sigma_x{}^2} = \frac{hk \ \text{Cov}(U, V)}{h^2 \sigma_u{}^2}$$

$$= \frac{k}{h} \cdot \frac{\text{Cov}(U\,, V)}{\sigma_u{}^2} = \frac{k}{h} b_{vu}$$

Similarly, we can show that

$$b_{xy} = \left(\frac{h}{k}\right) b_{uv}.$$

10.8. SOLVED EXAMPLES: -

Example 10.8.1. Obtain the equations of the lines of regression for the following data. Also obtain the estimate of X for $Y = 70$.

X:	65	66	67	68	69	70	72
Y:	67	68	65	68	72	69	71

Solution. Let $U = X - 68$ and $V = Y - 69$, then $\overline{\text{U}} = 0, \ \overline{\text{V}} = 0,$ $\sigma_U{}^2 = 4.5, \ \sigma_v{}^2 = 5.5$ $\text{Cov}(U, V) = 3$ and $r(U, V) = 0.6$.Since correlation coefficient is independent of change of origin, we get $r = r(X, Y) = r(U, V) = 0.6$.We know that If $U = \frac{X-a}{h}\,, V = \frac{Y-b}{k}$, then $\bar{X} = a + h\bar{U}, \ \bar{Y} = b + k\bar{V}, \ \sigma x = h\ \sigma y$ and $\sigma y = k \sigma y$
In our case $h = k = 1, \ a = 68$ and $b = 69$ $\sigma_X = \sigma_U = \sqrt{4.5} = 2.12$ and$\sigma_Y = \sigma_V = \sqrt{5.5} = 2.35$ Equation of line of regression Y on X is $Y = \bar{Y} = r \frac{\sigma_y}{\sigma_x} (X - \bar{X})$ $i.e.,$ $Y = 69 + 0.6 \times \frac{2.35}{2.12} (X - 68)$
Hence, $Y = 0.665X + 23.78$ Equation of line of regression of X on Y is $X - \bar{X} = r \frac{\sigma_x}{\sigma_y} (Y - \bar{Y})$ This gives, $X = 68 + 0.6 \times \frac{2.12}{2.35} (Y - 69)$ i. e. $X = 0.54Y + 30.74$.To estimate X for given Y,we use the line of regression of X on Y. If$Y = 70$, estimated value of X is given by $\hat{X} = 0.54 \times 70 + 30.74 = 68.54$.Where $\hat{X}$ is estimate of X,for$Y = 70$.

Example 10.8.2. In a partially destroyed laboratory record of an analysis of correlation data, the following results only are legible:
Variance of $\text{X} = 9$. Regression equations : $8\text{X} - 10\text{Y} + 66 = 0, \ 40\text{X} - 18\text{Y} = 214$. What were, (i) the mean value of X and Y (ii) the correlation coefficient between X and Y , and (iii) the standard deviation of Y?

Solution (i) Since both the lines of regression pass through the point $(\bar{X}, \bar{Y})$, we have $8\bar{X} - 10\bar{Y} + 66 = 0$, and $40\bar{X} - 18\bar{Y} = 214$. Solving these, we get $\bar{X} = 13, \bar{Y} = 17$.

(ii)Let $8X - 10Y + 66 = 0$and $40X - 18Y = 214$ be the lines of regression of Y on X and X and Y respectively. These equations can be put in the form:$Y = \frac{8}{10}X + \frac{66}{10}$and $X = \frac{18}{40}Y + \frac{214}{40}$

$\therefore$ $b_{yx} =$ Regression coefficient of Y on X $= \frac{8}{10} = \frac{4}{5}$

and $b_{xy} =$ Regression coefficient of X on $Y = \frac{18}{40} = \frac{9}{20}$

Hence $r^2 = \pm\frac{3}{5} = \pm 0.6$

But since both the regression coefficients are positive, we take $r = +0.6$

(iii)We have $b_{yx} = r\frac{\sigma_x}{\sigma_y}$

This implies that $\frac{4}{5} = \frac{3}{5} \times \frac{\sigma_y}{3}$. Hence $\sigma_y = 4$.

Remarks.10.8.1. (a) It can be verified that the values of $\bar{X} = 13$ and$\bar{Y} = 17$ as obtained in part $(i\ )$ satisfy both the regression equations. In numerical problems of this type, this check should invariably be applied to ascertain the correctness of the answer.

(b) If we had assumed that $8X - 10Y + 66 = 0$, is the equation of the line of regression of X on Y and $40X - 18Y = 214$ is the equation of line of regression of Y on X , then we get respectively: $8X = 10Y - 66$ and $18Y = 40X - 214$ i.e. $X = \frac{10}{8}Y - \frac{66}{8}$and $Y = \frac{40}{18}X - \frac{214}{18}$

After comparing we get, $b_{xy} = \frac{18}{8}$ and $b_{yx} = \frac{40}{18}$

$\therefore r^2 = b_{xy}.b_{yx} = \frac{10}{8} \times \frac{40}{18} = 2.78.$

But since r^2always lies between 0 and 1, our supposition is wrong.

Example 10.8.3. Find the most likely price in Bombay corresponding to the price of Rs. 70 at Calcutta from the following:

	Calcutta	Bombay
Average Price	65	67
Standard deviation	2.5	3.5

Correlation coefficient between the prices of commodities in the two cities is 0.8.

Solution. Let the prices (in Rupees), in Bombay and Calcutta be denoted by$Y$ and X respectively. Then we are given $\bar{X} = 65,\ \bar{Y} = 67, \sigma_x = 2.5,\ \sigma_y = 3.5$ and $r = r(X, Y) = 0.8,$ we want Y for $X = 70$. Thus, the line of regression of Y on X is: $Y - \bar{Y} = r.\frac{\sigma_y}{\sigma_x}(X - \bar{X})$ $Y = 67 + 0.8 \times \frac{3.5}{2.5}(X - 65)$. When $X = 70,$ $\hat{Y} = 67 + 0.8 \times \frac{3.5}{2.5}(70 - 65) = 72.6.$

Example 10.8.4. Can Y=5+2.8X and X=3-0.5Y be the estimated regression equations of Y on X and X on Y respectively? Explain your answer with suitable theoretical arguents.

Solution. Line of regression Y on X ,$Y = 5 + 2.8X$ gives $b_{yx} = 2.8$ .And line of regression of X on Y, $X = 3 - 0.5Y$ gives $b_{xy} = -0.5$. Which is not possible, since each of the regression coefficients b_{yx} and b_{xy} must have the same sign.

Example 10.8.5. Given

$$f(x,y) = xe^{-x(y+1)}; x \geq 0, y \geq 0,$$

Find the regression curve of Y on X.

Solution. Marginal p.d.f. of X is given by$f_1(x) = \int_0^\infty f(x,y)dy = \int_0^\infty xe^{-x(y+1)}\,dy = xe^{-x}\int_0^\infty e^{-xy}\,dy = xe^{-x}\left|\frac{e^{-xy}}{-x}\right|_0^\infty = e^{-x}, x \geq 0$

Conditional p.d.f. of Y on X is given by

$$f(y|x) = \frac{f(x,y)}{f_1(x)} = \frac{xe^{-x(y+1)}}{e^{-x}} = xe^{-xy}, y \geq 0.$$

The regression cure of Y on X is given by

$$y = E(Y|X = x) = \int_0^\infty y\, f(y|x)dy = \int_0^\infty yxe^{-xy}\,dy$$

$$= x[\left|\frac{ye^{-x}}{-x}\right|_0^\infty + \int_0^\infty \frac{e^{-xy}}{-x}dy\ ] = 0 + \left|\frac{e^{-xy}}{-x}\right|_0^\infty = \frac{1}{x}$$

$i.e.,$ $y = \frac{1}{x}$. Which is same as $xy = 1$. Which is the equation of a rectangular hyperbola. Hence the regression of Y on X is not linear.

Example 10.8.6. Obtain the regression equation of Y on X for the following distribution :

$$f(x,y) = \frac{y}{(1+x)^4}\exp\left(-\frac{y}{1+x}\right);\ x,y \geq 0$$

Solution. Marginal p.d.f. of X is given by

$$f_1(x) = \int_0^\infty f(x,y)dy = \frac{1}{(1+x)^4}\int_0^\infty ye^{-\frac{y}{1+x}}dy$$

$$= \frac{1}{(1+x)^4}\,.\,\Gamma 2\,.\,(1+x)^2$$

$$= \frac{1}{(1+x)^2}; x \geq 0$$

The conditional p.d.f. of Y(for given X) is

$$y = E(Y|X) = \int_0^\infty yf(y|x)dy = \frac{1}{(1+x)^2}\exp\left(-\frac{y}{1+x}\right)\ ; y \geq 0$$

Regression equation of Y on X is given by

$$y = E(Y|X) = \int_0^\infty yf(y|x)\,dy = \frac{1}{(1+x)^2}\int_0^\infty y^2e^{-\frac{y}{1+x}}dx$$

$$= \frac{1}{(1+x)^2}\,.\,\Gamma 3\,.\,(1+x)^3$$

$y = 2\,(1+x)$ $[\because \Gamma 3 = 2! = 2]$.

Hence the regression of Y on X is linear.

Example 10.8.7. *Let (X,Y) have the joint p.d.f. given by*

$$f(x,y)=\begin{cases}1, if\ |y|<x, 0<x<1\\ 0, otherwise\end{cases}$$

Show that the regression of Y on X is linear but regression of X on Y is not linear.

Solution:

$$f_1(x)=\int_{-x}^{x} f(x,y)dy=\int_{-x}^{x}1.dy=2x;\ \ 0<x<1$$

$$f_2(y)=\int_{|y|}^{1} f(x,y)dx=\int_{|y|}^{1}1.dx=1-|y|;\ -1<y<1$$

$$\therefore f_1(x|y)=\frac{f(x,y)}{f_2(y)}=\frac{1}{1-|y|};\ \ -1<y<1,\ \ 0<x<1$$

$$=\begin{cases}\dfrac{1}{1-y}, 0\ <y<1<x<1\\ \dfrac{1}{1+y}, -1\ <y<0<x<1\end{cases}$$

$$f_2(x|y)=\frac{f(x,y)}{f_1(x)}=\frac{1}{2x}\ , 0<x<1; |y|<x$$

$$E(Y|X=x)=\int_{-x}^{x} y.f_2(y|x)dy=\int_{-x}^{x}\frac{y}{2x}dy=\frac{1}{4x}.|y^2|_{-x}^{\ x}=0$$

Hence the curve of regression of Y on X is $y=0$, which is a straight line.

$$E(X|Y=y)=\int x\, f_1(x|y)\, dx$$

$$\therefore\ \ E(X|Y=y)=\int_0^1 x\left(\frac{1}{1-y}\right)dx=\frac{1}{2(1-y)}, 0<y<1$$

And $\qquad E(X|Y=y)=\int_0^1 x\left(\frac{1}{1+y}\right)dx=\frac{1}{2(1+y)}, -1<y<0$

Hence the curve of regression of X on Y is

$$x=\begin{cases}\dfrac{1}{2(1-y)}, 0<y<1\\ \dfrac{1}{2(1+y)},\ \ -1<y<0,\end{cases}$$

Which is not a straight line.

Example 10.8.8. Variables X on Y have the joint p.d.f.

$$\mathrm{f(x,y)}=\frac{1}{3}(\mathrm{x+y}),\qquad 0\le \mathrm{x}\le 1,\qquad 0\le \mathrm{y}\le 2.$$

Find:

(i) r(X, Y)
(ii) The two lines of regression
(iii) The two regression curves for the means.

Solution: The marginal p.d.f.'s of X and Y are given by :

$$f_1(x) = \int_{-x}^{x} f(x,y)dy = \frac{1}{3}\int_0^2 (x+y)dy = \frac{1}{3}\left|xy + \frac{y^2}{2}\right|_0^2$$

i.e. $f_1(x) = \frac{2}{3}(1+x);\ 0 \le x \le 1$

$$f_2(y) = \int_0^1 f(x,y)\,dx = \frac{1}{3}\int_0^1 (x+y)dx = \frac{1}{3}\left|\frac{x^2}{2} + xy\right|_0^1$$

i.e. $f_2(y) = \frac{1}{3}\left(\frac{1}{2} + y\right);\ 0 \le y \le 2$

The conditional distributions are given by:

$$f_3(y|x) = \frac{f(x,y)}{f_1(x)} = \frac{1}{2}\left(\frac{x+y}{1+x}\right)$$

$$f_4(x|y) = \frac{f(x,y)}{f_2(y)} = \frac{2(x+y)}{1+2y}$$

$$E(Y|x) = \int_0^2 y.f_3(y|x)\,dy = \frac{1}{2(1+x)}\int_0^2 y(x+y)dy$$

$$= \frac{1}{2(1+x)}\left|\frac{xy^2}{2} + \frac{y^3}{3}\right|_{y=0}^{y=2} = \frac{3x+4}{3(x+1)}$$

Similarly, we shall get

$$E(X|y) = \int_0^1 x\,f_4(x|y)dx = \frac{2}{1+2y}\int_0^1 (x^2+xy)dx = \frac{2+3y}{3(1+2y)}$$

(iii) Hence the regression curves for means are:

$$y = E(Y|x) = \frac{3x+4}{3(x+1)} \text{ and } \quad x = E(X|y) = \frac{2+3y}{3(1+2y)}$$

From the marginal distributions we shall get

$$E(X) = \int_0^1 x\,f_1(x)dx = \frac{5}{9}, \quad E(X^2) = \int_0^1 x^2\,f_1(x)dx = \frac{7}{18}$$

$Var(X) = {\sigma_x}^2 = \frac{7}{18} - \left(\frac{5}{9}\right)^2 = \frac{13}{162}$.Similarly, we shall get$E(Y) = \frac{11}{9},\ E(Y^2) = \frac{16}{9}\ ,\ {\sigma_y}^2 = \frac{16}{9} - \left(\frac{11}{9}\right)^2 = \frac{23}{81}$. Also $E(XY) = \int_0^1\int_0^2 xy\,f(x,y)dxdy = \frac{1}{3}\int_0^1\int_0^2 (x^2y + xy^2)dxdy$

$$= \frac{1}{3}\left\{\left(\int_0^1 x^2dx\right)\left(\int_0^2 y\,dy\right) + \left(\int_0^1 x\,dx\right)\left(\int_0^2 y^2\,dy\right)\right\}$$

$$= \frac{1}{3}\left[\frac{1}{3}\times 2 + \frac{1}{2}\times\frac{8}{3}\right] = \frac{2}{3}$$

$\therefore$ $\quad \text{Cov}(X,Y) = E(XY) - E(X)\,E(Y) = \frac{2}{3} - \frac{5}{9}\times\frac{11}{9} = -\frac{1}{81}$

(i) $\quad r(X,Y) = \frac{\text{Cov }(X,Y)}{\sigma_x.\sigma_y} = \frac{-\frac{1}{81}}{\sqrt{\frac{13}{162}\times\frac{23}{81}}} = -\left(\frac{2}{299}\right)^{\frac{1}{2}}$

(ii) The two lines of regression are:

$$Y - E(Y) = \frac{\text{Cov }(X,Y)}{\sigma_x.\sigma_y} = [X - E(X)] \quad \Rightarrow \quad Y - \frac{11}{9}$$

$= -\frac{2}{13}\left(X - \frac{5}{9}\right)$ and $X - E(X) = \frac{\text{Cov }(X,Y)}{\sigma_y{}^2}[Y - E(Y)] \Rightarrow X - \frac{5}{9} = -\frac{1}{23}\left(Y - \frac{11}{9}\right)$

CHECK YOUR PROGRESS

Q.1. When the curve is a straight line then the regression is said to be linear. True/False

Q.2 If r = 0 then two line of regression is parallel to axes. True/False

Q.3 Coefficient of correlation is the geometric mean of the coefficient of regression.
True/False

Q.4 The regression coefficient of x on y is 3.2 and that of y on x is 0.8. True/False

Q.5 Regression coefficient are independent of the change of origin but not of ……….

Q.6 Arithmetic mean of coefficient of regression is greater than the coefficient of……….

Q.7 The regression line of y on x from the following data is…...

X	1	2	3	4	5	6	7	8	9	10
Y	10	12	16	28	25	36	41	40	42	50

Q.8 The product of regression coefficients is less than or equal to.
a. 0 b. 1 c. 2 d. -1

Q.9 If $b_{yx}>1$, then $b_{xy} \le 1$ provided $b_{yx}b_{xy} \le$
a. 1 b. 2 c. 0 d. none

Q.10 If r = 1 or -1 then the two regression lines will
a. parallel b. perpendicular c. Coincide
d. Intersecting

10.9 SUMMARY: -

In this unit, we have studied the regression, lines of regression. We have also studied, how the regression coefficients give important information in context of line of regression and regression coefficients.

10.10 GLOSSARY:-

(i) Line of regression
(ii) Regression coefficients.
(iii) Residual Variance

10.11.REFERENCE BOOKS:-

1. S. C. Gupta and V. K. Kapoor, (2020), *Fundamentals of mathematical statistics*, Sultan Chand & Sons.
2. Seymour Lipschutz and John J. Schiller, (2017), *Schaum's Outline: Introduction to Probability and Statistics*, McGraw Hill Professional.
3. J. S. Milton and J. C. Arnold , (2003*), Introduction to Probability and Statistics (4th Edition),* Tata McGraw-Hill.
4. https://www.wikipedia.org.

10.12.SUGGESTED READINGS:-

1. A.M. Goon,(1998), *Fundamental of Statistics 7th Edition),* 1998.
2. R.V. Hogg and A.T. Craig, (2002), *Introduction to Mathematical Statistics,* MacMacMillan, 2002.
3. Jim Pitman, (1993), *Probability,* Springer-Verlag.
4. https://archive.nptel.ac.in/courses/111/105/111105090.

10.13 TERMINAL QUESTIONS:-

TQ 10.13.1 Explain what are regression lines. Why are there two such lines? Also derive their equations.

TQ 10.13.2 Define, line of regression and regression coefficient. Also find th equations to the lines of regression and show that the coefficient of correlation is the geometric mean of coefficients of regression.

TQ 10.13.3 Obtain the equation of the line of regression of Y on X and show that
the angle θ, between the two lines of regression is given by:

$$\tan(\theta) = \frac{1-\rho^2}{\rho} \times \frac{\sigma_1\sigma_2}{{\sigma_1}^2 + {\sigma_2}^2}$$

TQ 10.13.4 If θ is the acute angle between the two regression lines with correlation coefficient r, then show that $\sin(\theta) = 1 - r^2$.

TQ 10.13.5 For the following data obtain the equations of the lines of regression and obtain an estimate of Y which should correspond to X = 6.2.

X	1	2	3	4	5	6	7	8	9
Y	9	8	10	12	11	13	14	16	15

10.14. ANSWER:-

ANSWER OF CHECK YOUR PROGRESS

CHQ.1 TRUE
CHQ.2 TRUE
CHQ.3 TRUE
CHQ.4 FALSE
CHQ.5 SCALE
CHQ.6 CORELATION
CHQ .7 $Y - 30 = 4.48\,(X - 5.5)$
CHQ.8 B
CHQ.9 A
CHQ.10 C

ANSWER OF TERMINAL QUESTIONS

TQ10.13.5 $Y - 12 = 0.95(X - 5)$ & $X - 5 = 0.95(Y - 12)$ and correspond to $X = 6.2$ estimated value of Y is: 13.14

BLOCK V
CONCEPT OF STATISTICAL HYPOTHESIS

UNIT 11:- BASICS OF SAMPLING

CONTENTS:

11.1.INTRODUCTION:-

In this unit we are explaining about basics of sampling theory. Suppose we want to analyse some statistical properties from a large number of individuals, items or things, like from: population of a country, all tree in a forest, street dog in a state and so on. It is very impractical that we investigate and analyse all the individuals, items or things. The group of these specific individuals, items or things under study is called population. Hence complete and exact analysis of population is very tough task. To deal this we need sampling theory. In statistics, quality assurance, and survey methodology, sampling is the selection of a subset (a statistical sample) of individuals from within a statistical population to estimate characteristics of the whole population. Statisticians attempt to collect samples that are representative of the population in question. Sampling has lower costs and faster data collection than measuring the entire population and can provide insights in cases where it is infeasible to measure an entire population.

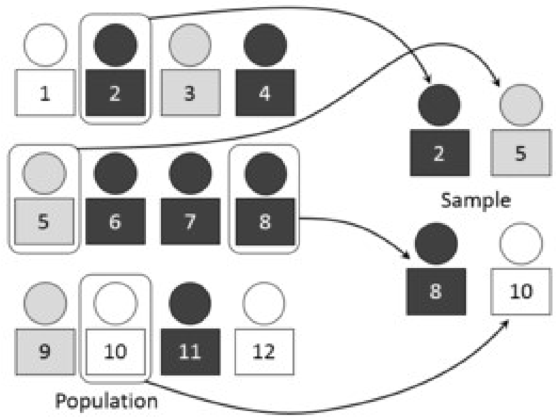


Fig. 11.1.1
Ref: https://en.wikipedia.org/wiki/Sampling_(statistics)
A visual representation of the sampling process

11.2 OBJECTIVES:-

After studying this unit learner will be able to:

1. Differentiate the types of sample drawn.
2. Understand the statistic and parameter.
3. Approximate some basic constant of population.

11.3.POPULATION AND SAMPLE:-

Population: A population is a group or set of individuals, items or things, which are under the study of statistical investigation. Note that population may be infinite set. An element of a population is called statistical individual of that population.

Example 11.3.1 Total students enrolled in higher education in a country, Total soap produced in a soap company during a particular month etc.

Sample: A finite subset of a population of is called sample of that population. Total number of statistical individuals in the sample is called sample size. The elements in sample is called sample unit. Some examples of sample of population related to example 1 are as follows:

Example 11.3.2 Picking a student randomly from each university, colleges and institutes. This group is a sample of total students enrolled in higher education. Now for the later one, pick five soaps produced everyday, at the last day of the month all chosen soaps on the daily basis form a sample of mentioned population.

Problem 11.3.3 Whatare the advantages of sampling?

Solution: To investigate measure of the population by a sample has many advantages. Some of the advantages are as following: Time saving, Cost saving, Easy to conduct and Desirable. If we do investigation on every individual of the population, then it will take more time and cost. Also,an investigation on less number of units are more ease to conduct. Now, suppose we want to check lifeline of bulb produced in a company, then we cannot check each bulb. Otherwise, there is no bulb available for sell. So, sometime checking the whole population units are not desirable.

11.4.TYPES OF SAMPLING:-

In the context of picking process, we have four types of sampling, which are as follows:

(i) Purposive sampling
(ii) Random sampling
(iii) Stratified sampling
(iv) Systematic sampling

Purposive Sampling: If the selected sample unit has some pre purpose of being selected in the sample. Then this type of sample is called purposive sample. For example, if a teacher wants to show his/ her class students excellence by a sample student, he/ she will chose brilliant student of that class. This type of sampling is biased in nature and approximation of properties of population from the purposive sampling has much more error.

Problem 11.4.1 What is the key drawback of purposive sampling.

Solution: The key drawback of purposive sampling is favouritism and nepotism.

Random Sampling: In the random sampling, sample units are selected at random. In this case drawback of purposive sampling is totally overcome. Thus, in random sample, each unit of population has an equal chance of being selected in sample. Let the size of population\ is N and we want to draw a sample of size n, then there are: $\binom{N}{n}$ possible samples. One of the most commonly used examples of random sampling is lottery system.

Problem 11.4.2 What are the drawbacks of random sampling.

Solution: If the population is heterogeneous, then in this sample we may have missing of sample representative of some homogenous group of population. For example, if we want a random sample of 10 students from co-education institute. Then chosen random sample may have only boy students. Thus, the girl student's information is totally missing in this sample.

Simple Random Sampling: Simple random sampling is type of random sampling in which each unit of the population has an equal chance of being selected in sample, with an extra condition that this probability is independent of previous drawing. If the population size is finite then simple sampling must be done with replacement and if the population is infinite then replacement is not required.

Stratified Sampling: In this type of sampling, the units of population are heterogenous. We divide this population in homogenous groups. These groups are called strata. Each strata differ from each other, but is homogenous within itself. Then units of sample is chosen randomly from each of these groups. The number of units selected in each groups varies according to the statistical investigation and its relative importance. The resultant sample is termed as stratified sample. This sample is best representative of the population.

Systematic Sampling: In this sampling, we choose first sample unit at random from the population. And after that all the sample unit are chosen by some definite rule. For example, suppose we have list of students in M.Sc. Mathematics and we want a sample from this list. Then choose first student at random and then 4^{th} student from previously selected student may be chosen to form a sample. This type of sample is known as systematic.

Remark11.4.3: In the context of probability, sample is classified as following:

(i) **Probability Sampling:** If every individual of population has a chance of being selected in the sample. Then the sampling is called probability sampling. Random sampling, stratified sampling, systematic sampling are probability sampling.

(ii) **Non- probability Sampling:** If some of the individuals of population have almost possibility of being selected and some of the individuals do not have the chance of being selected in the sample. This type of sample is biased and involve purpose in selection process. Example of non- probability sampling is purposive sampling.

(iii) **Mixed Sampling:** In this sampling, both the probability sampling and non-probability sampling are involved. Suppose we want to chose a sample of 10 students for a quiz and 6 students are chosen random while 4 brilliant students are chosen. Then this is an example of mixed sampling.

11.5. STATISTIC AND PARAMETER:-

Any statistical measure based on population data is known as parameter and any statistical constant based on sample is known as statistic. For e.g. mean of population is a parameter and mean of sample is a statistic. In the following table we fix parameter and statistic symbol for mean and variance:

S.N.	Statistical measure	Parameter Symbol Θ	Statistic Symbol t
1	Mean	μ	$\bar{x}$
2	Variance	σ^2	S^2

Consider a sample of size n. Let $x_1, x_2, \ldots, x_n$ be sample values. Note that each $x_i's$ are random variable, which can take any individual values of population. Thus distribution of each$x_i's$ are identical and same as distribution of r.v. of population. Further a statistic $t = t(x_1, x_2, \ldots, x_n)$ is a function of sample values $x_1, x_2, \ldots, x_n$. Clearly a statistic t is also a random variable. A statistic is termed as estimate of its parameter. For e.g. sample mean is estimate of population mean. The sample mean $(\bar{x})$is defined as:$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$.

Unbiased Estimate: A statistic t is known as unbiased estimate of Θ if $E(\text{t}) = \Theta$. $E(\text{Statistic}) = Parameter \ldots\ldots\ldots\ldots\ldots\ldots (\mathbf{11.5.1})$

Remark 11.5.1 Throughout we fix s^2 (little s square) for sample variance. Hences^2 is defined as: $s^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}$.
But we take S^2 as an estimate of population varianceσ^2, which is defined as follows: $S^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}$.
The reason for this is that S^2 is an unbiased estimator of population variance σ^2. More detail about this will be discuss in the next unit.

Standard Error: As we know a statistic t is a random variable. More precisely if we have population size N and sample size n, then we have $\binom{N}{n} = m(say)$ possible samples. i.e. in this case random variable t can be any one of the m values, $t_1, t_2, \ldots, t_m$. Now standard deviation of this t is known as standard error of statistic t. Standard error of statistic t is denoted as SE(t).
Remark 11.5.2: If sample size n is large, then for any statistic$t$:$Z = \frac{t-E(t)}{SE(t)} \sim N(0,1)$.

Check Your Progress

Problem 1: The number of possible sample of size n out of N population units without replacement is : ………..

Problem 2: The number of possible sample of size n out of N population units with replacement is :………..

Problem 3: An unordered sample of size n can occur in how many ways:….

Problem 4: Probability of any one sample of size n being drawn out of N units is………..

Problem 5: A function of variates for estimating a parameter is called:……

Problem 6: If the observations recorded on five sample are 3,4,5,6,7, then the sample variance is: ………..

Problem 7: A sample of size n is drawn from a dichotomous population. If the sample has proportion p of items of category I and q of category II, then the variance of the proportion p is: ………..

Problem 8: If the sample values are 1,3,5,7,9, then the standard error of sample mean is: ………..

Problem 9: If n units are selected in a sample from N population units, then the sampling fraction is: ………..

Problem 10: A sample of 16 items from an infinite population having S.D.= 4, yielded total scores as 160. The standard error of sampling distribution of mean is:………..

11.6. HYPOTHESIS, ERROR, LEVEL OF SIGNIFICANCE AND PROCEDURE:-

Tests of Significance. In sampling theory we study tests of significance, which involve some theoretical statistical concept to give some conclusion about the sample. Suppose we have a sample data, then on this data many questions arises like:

- There is no significance difference between the observed sample statistic and the population hypothetical parameter.

- The obtained sample data, fit significantly to the population. Here significance means, allowing some percentage of error.

Remark 11.6.1: For large n, most of the distribution like: Binomial, Poisson, t-distribution, Chi- square distribution, F- distribution follows normal distribution. Therefore for large sample we use Normal test of significance. And for small sample we do test of significance using t-test. F-test and Fisher's z-transformation.

Null Hypothesis. A hypothesis is the statement that is not known that it is true or not. Under the test of significance, we first make null hypothesis. The null hypothesis is hypothesis which is independent, hypothesis of no difference. For example, if we want to make a statement about to compare average performance of class A and class B. So we have three choice (i) average performance of class A is same as of class B
(ii) average performance of class A is better than class B (iii) average performance of class B is better than class A. Out of these three statement, statement (i) will be the null hypothesis, because statement (ii) & (iii) showing some biasedness and difference view on class A and class B. The null hypothesis is denoted by H_0.

Alternative Hypothesis. Any hypothesis which is complementary (negation or part of negation) to the null hypothesis is called an alternative hypothesis. It is denoted by H_1. Note that there is no meaning of alternative hypothesis, till the null hypothesis is not define.

Example: Suppose we have sample and we want to test that the population mean μ is same as μ_0 or not at some level of significance. Then null hypothesis H_0 should be, $H_0: \mu = \mu_0$. Then the alternative hypothesis can be one of the following:

(i) $H_1: \mu \neq \mu_0$
(ii) $H_1: \mu > \mu_0$
(iii) $H_1: \mu < \mu_0$

The alternative hypothesis in (i) is known as a two tailed alternative and the alternatives in (ii) and (iii) are known as right tailed and left-tailed alternatives respectively. The setting of alternative hypothesis is very important since it enables us to decide whether we have to use a single-tailed (right or left) or two-tailed test.

Errors in Sampling. The main objective in sampling theory is to draw valid inferences about the population parameters on the basis, of the sample results. In practice we decide to accept or reject the lot after examining a sample from it. As such we are liable to commit the following two types of errors:

Type I Error : Reject H_0 when it is true.

Type II Error : Accept H_0 when it is wrong, i.e., accept H_0 when H_1 is true. If we write,

$$\text{P(Reject H}_0\text{ when it is true)} = \text{P (Reject H}_0|\text{H}_0) = \alpha$$

and
P (Accept H_0 when it is wrong) $=$ P(Accept $H_0|H_1$) $= \beta$**...(11.6.1**)
then α and β are called the sizes of type I error and type II error, respectively.

In practice, type I error amounts to rejecting a lot when it is good and type II error may be regarded as accepting the lot when it is bad.
Thus P(Reject a lot when it is good) $= \alpha$
And P (Accept a lot when it is bad) $= \beta$…**.........(11.6.2)**
where α and β are referred to as Producer's risk and Consumer's risk respectively.

Critical Region and Level of Significance: A region (corresponding to a statistic t) in the sample space S which amounts to rejection of H_0 is termed as critical region or region of rejection: If ω is the critical region and if $t = t(x_1, x_2, \dots, x_n)$ is the value of the statistic based on a random sample of size n, then

$$P(t\epsilon\omega|H_0) = \alpha, P(t\epsilon\overline{\omega}|H_1) = \beta \text{..............}\textbf{(11.6.3)}$$

where $\overline{\omega}$ is the complementary set of ω. is called the acceptance region. We have $\omega \cup \overline{\omega} = S \; and \; \omega \cap \overline{\omega} = \Phi$.

The probability α that a random value of the statistic t belongs to the critical region is known as the level of significance. In other words, level of significance is the size of the type I error (or the maximum producer’s risk). The levels of significance usually employed in testing of hypothesis are 5% or 1%. The level of significance is always fixed in advance before collecting the sample information.

One tailed and Two Tailed Tests. In any test, the critical region is represented by a portion of the area under the probability curve of the sampling distribution of the test statistic.

A test of any statistical hypothesis where the alternative hypothesis is one tailed (right tailed or left tailed) is called a one tailed test. For example, a test for testing the mean of a population, $H_0: \mu = \mu_0$ against the alternative hypothesis:
$H_1: \mu > \mu_0$. (Right tailed) or $H_1: \mu < \mu_0$ (Left tailed) is a single tailed test.

In the right tailed test ($H_1: \mu > \mu_0$), the critical region lies entirely in the right tail of the sampling distribution of sample mean, while for the left tail test ($H_1: \mu < \mu_0$), the critical region is entirely in the left tail of the distribution. A test of statistical hypothesis where the alternative hypothesis is two tailed such as:

$H_0: \mu = \mu_0$ against the alternative hypothesis $H_1: \mu \neq \mu_0$, ($H_1: \mu > \mu_0$ or $H_1: \mu < \mu_0$), is known as two tailed test and in such a case the critical region is given by the portion of the area lying in both the tails of the probability curve of the test statistic. In a particular problem, whether one tailed or two tailed test is to be applied depends entirely on the nature of the alternative hypothesis. If the alternative hypothesis is two-tailed we apply two-tailed test and if alternative hypothesis is one-tailed, we apply one tailed test. For example, suppose that there are two population brands of bulbs, one manufactured by standard process (with mean life μ_1) and the other manufactured by some new technique (with mean life μ_2).

If we want to test if the bulbs differ significantly, then our null hypothesis is $H_0: \mu_1 = \mu_2$ and alternative will, be $H_1: \mu_1 \neq \mu_2$thus giving us a two-tailed test. However, if we want to test if the bulbs produced by new process have high average life than those produced by standard process, then we have $H_0: \mu_1 = \mu_2$ and $H_1: \mu_1 < \mu_2$, thus giving us a left-tail test Similarly, for testing if the product of new process is inferior to that of standard process, then we have $H_0: \mu_1 = \mu_2$ and $H_1: \mu_1 > \mu_2$. thus, giving us a right-tail test. Thus, the decision about applying a two-tail test or one-tail (right or left) test will depend on the problem under study.

Critical Values or Significant Values The value of test statistic which separates the critical (or rejection) region and the acceptance region is called the critical value or significant value. It depends upon:
(i) The level of significance used, and
(ii) The alternative hypothesis, whether it is two-tailed or single-tailed.
As has been pointed earlier for large samples, the standardized variable corresponding to the statistic $t$viz.,

$$Z = \frac{t - E(t)}{S.E(t)} \sim N(0,1), \dots\dots\dots\dots\dots\dots. (\mathbf{11.6.4})$$

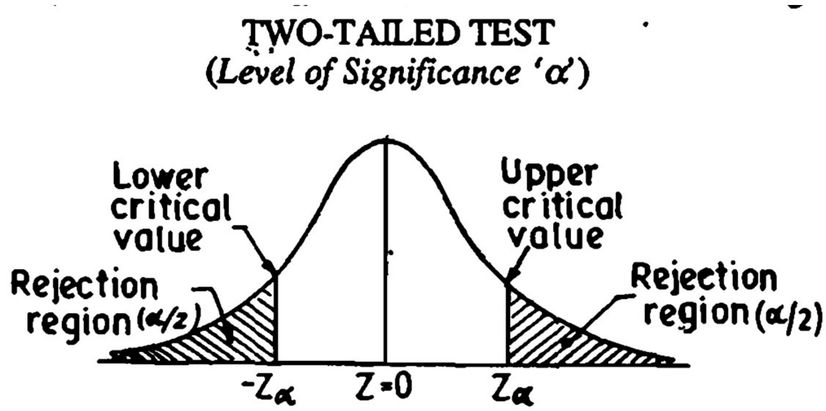
which is asymptotically as $n \to \infty$. The value of Z given by (11.6.4) under the null hypothesis is known as test statistic. The critical value of the test statistic at level of significance α for a two-tailed test is given by z_α where z_α, is determined by the equation

$$P(|Z| > z_\alpha) = \alpha \dots\dots\dots\dots\dots\dots\dots\dots\dots (\mathbf{11.6.5})$$

i.e., z_α is the value so that the total area of the critical region on both tails is α.
Since normal probability curve is a symmetrical curve, from $(\mathbf{11.6.5})$, we get

$$P(Z > z_\alpha) + P(Z < -z_\alpha) = \alpha$$
$$\Rightarrow P(Z > z_\alpha) + P(Z > z_\alpha) = \alpha$$
$$\Rightarrow P(Z > z_\alpha) = \alpha/2.$$

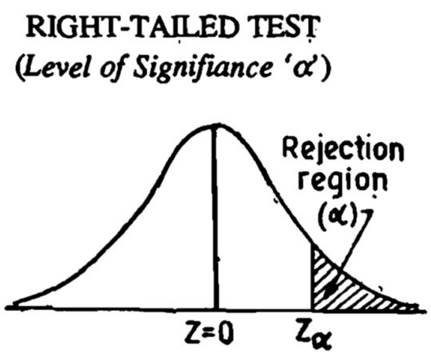
i.e. the area of each tail is $\alpha/2$. Thus z_α is the value such that area to the right of z_α is $\alpha/2$ and, to the left of $-z_\alpha$ is $\alpha/2$, as shown in the following diagram.

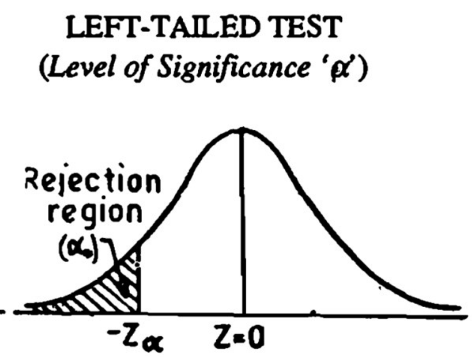


In case of single-tail alternative, the critical value z_α is determined so that total area to the right of it (for right-tailed test) is α and for left-tailed test the total area to the left of $-z_\alpha$is α. i.e.

For Right-tail Test: $P(Z > z_\alpha) = \alpha$ ... ... $(\mathbf{11.6.6})$

And, for Left-tail Test: $P(Z < -z_\alpha) = \alpha$ ... ... $(\mathbf{11.6.7})$





Thus the significant or critical value of Z for a single-tailed test (left,or right) at level of significance α is same as the critical value of Z for a two-tailed test at level of significance 2α. Consider the following table.

Level of significance (α)	**1%**	**5%**	**10%**
Critical values (z_α) in two tailed test	$\|z_\alpha\|$=2.58	$\|z_\alpha\|$=1.96	$\|z_\alpha\|$=1.645
Critical values (z_α) in right tailed test	z_α=2.33	z_α=1.645	z_α=1.28
Critical values (z_α) in left tailed test	z_α=-2.33	z_α=-1.645	z_α=-1.28

Remark11.6.2. If *n* is small, then the sampling distribution of the test statistic Z will not be normal and in that case we can't use the above significant values, which have been obtained from normal probability curves. In this case, viz., *n* small, (usually less than 30), we use the significant values based on the exact sampling distribution of the statistic Z. which turns out to be t-test. F-test and Fisher's z-

transformation. These significant values have been tabulated for different values of n and α.

Procedure for Testing of Hypothesis. We now summarize below the various steps in testing of a statistical hypothesis in a systematic manner.

1. Null Hypothesis. Set up the Null Hypothesis H_0.
2. Alternative Hypothesis. Set up the Alternative Hypothesis H_1.
3. Level of Significance. Choose the appropriate level of significance α depending on the reliability of the estimates and permissible risk. This is to be decided before sample is drawn, i.e..α is fixed in advance.
4. Test Statistic (or Test Criterion). Compute the test statistic, under the null hypothesis

$$Z = \frac{t - E(t)}{S.E.(t)}$$

5. Conclusion. We compare the computed value of Z in step 4 with the significant value (tabulated value) z_α, at the given level of significance, α. If $|Z| < z_\alpha$, i.e. if the calculated value of Z (in modulus value) is less than z_α we say it is not significant. And we accept the null hypothesis at level of significance, α. And if $|Z| > z_\alpha$, then we say that it is significant and the null hypothesis is rejected at level of significance, α.

Test of Significance for Single Mean. Consider a population which follows normal distribution and $x_i, (i = 1,2, \ldots, n)$ is a random sample of size n then $\overline{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$. Furthermore if the sample size n is large and population does not follow normal distribution, then in this case also, by Central Limit Theorem, sample mean is distributed normally with mean μ and variance $\frac{\sigma^2}{n}$. Hence, for large samples, the standard normal variate corresponding to $\overline{x}$ is given by:

$$Z = \frac{\overline{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.(\mathbf{11.6.8})$$

Hence for large samples the null hypothesis H_0 is: there is no significant difference between the sample mean $(\overline{x})$ and population mean (μ).

Remarks 11.6.3 If the population standard deviation σ is unknown then we use its estimate. Since for large sample, $\widehat{\sigma^2} \simeq s^2$ therefore $\hat{\sigma} \simeq s$.

2. Confidence limits for μ. 95% confidence interval for μ is given by :

$$|Z| \le 1.96, ie., \left|\frac{\overline{x}-\mu}{\frac{\sigma}{\sqrt{n}}}\right| \le 1.96$$

$$\Rightarrow \quad \overline{x} - 1.96\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{x} + 1.96\frac{\sigma}{\sqrt{n}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.(\mathbf{11.6.9})$$

and $\overline{x} \pm 1.96\frac{\sigma}{\sqrt{n}}$are known as 95% confidence limits for μ. Similarly, 99% confidence limits for μ are and 98% confidence limits for are $\overline{x} \pm 2.33\frac{\sigma}{\sqrt{n}}$.However, in sampling from a finite population of size N , the corresponding 95% and 99% confidence limits for μ are respectively $\overline{x} \pm 1.96\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$ and $\overline{x} \pm 2.58\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$.

Test of Significance for Difference of Means. Let $\overline{x_1}$ be the mean of a random sample of size n_1 from a population with mean μ_1 and variance σ_1^2 and let $\overline{x_2}$ be the mean of an independent random sample of size n_2 from another population with mean μ_2 and variance σ_2^2 . Then, since sample sizes are large,

$$\overline{x_1} \sim N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right) \ and \ \overline{x_2} \sim N\left(\mu_2, \frac{\sigma_2^2}{n_2}\right)$$

Also $\overline{x_1} - \overline{x_2}$, being the difference of two independent normal variates is also a normal variate. The $Z(S.N.V)$ corresponding to $\overline{x_1} - \overline{x_2}$ is given by

$$Z = \frac{(\overline{x_1} - \overline{x_2}) - E(\overline{x_1} - \overline{x_2})}{S.E.(\overline{x_1} - \overline{x_2})} \sim N(0,1) \ldots\ldots\ldots\ldots\ldots\ldots(\mathbf{11.6.10})$$

Under the null hypothesis $H_0: \mu_1 = \mu_2$. I.e. there is no significant difference between the sample means, we get

$$E(\overline{x_1} - \overline{x_2}) = E(\overline{x_1}) - E(\overline{x_2}) = \mu_1 - \mu_2 = 0; \ \ldots\ldots\ldots(\mathbf{11.6.11})$$

$$V(\overline{x_1} - \overline{x_2}) = V(\overline{x_1}) + V(\overline{x_2}) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}, \ldots\ldots\ldots\ldots(\mathbf{11.6.12})$$

The covariance term vanishes, since the sample means $\overline{x_1}$ *and* $\overline{x_2}$ are independent.

Thus under $H_0: \mu_1 = \mu_2$, the test statistic becomes (for large samples),

$$Z = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\left(\left(\frac{\sigma_1^2}{n_1}\right) + \left(\frac{\sigma_2^2}{n_2}\right)\right)}} \sim N(0,\ 1) \ldots\ldots\ldots\ldots\ldots\ldots(\mathbf{11.6.13})$$

Remarks 11.6.4 If $\sigma_1^2 = \sigma_2^2 = \sigma^2$, I.e. if the samples have been drawn from the populations with common S.D. σ, then under $H_0:\mu_1 = \mu_2$,

$$Z = \frac{\overline{x_1} - \overline{x_2}}{\sigma\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0,\ 1) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.(\mathbf{11.6.14})$$

Remarks 11.6.5 If in **(11.6.14)** ,σ is not known, then its estimate based on the sample variances is used. If the sample sizes are not sufficiently large, then an unbiased estimate of σ^2 is given by

$$\widehat{\sigma_2} = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{(n_1 + n_2 - 2)}$$

since $$E(\sigma_2) = \frac{(n_1-1)E(S_1^2)+(n_2-1)E(S_2^2)}{(n_1+n_2-2)}$$

$$= \frac{(n_1 - 1)\sigma^2 + (n_2 - 1)\sigma^2}{(n_1 + n_2 - 2)} = \sigma^2$$

But since sample sizes are large, $S_1^2 \approx s_1^2$, $S_2^2 \approx s_2^2$, $n_1 - 1 \approx n_1$, $n_2 - 1 \approx n_2$. therefore in practice, for large samples, the following estimate of σ^2 without any serious error is used :

$$\widehat{\sigma^2} = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2}$$

However, if sample sizes are small, then a small sample test, t-test for difference of means is to be used.

Remarks 11.6.6. If $\sigma_1^2 \neq \sigma_2^2$ and σ_1 *and* σ_2 are not known, then they are estimated from sample values. This results in some error, which is practically immaterial, if samples are very large. These estimates for large samples are given by

$$\widehat{\sigma_1^2} = S_1^2 \approx s_1^2$$
$$\widehat{\sigma_2^2} = S_2^2 \approx s_2^2$$

In this case, $(\mathbf{11.6.13})$ gives$Z = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\left(\left(\frac{s_1^2}{n_1}\right)+\left(\frac{s_2^2}{n_2}\right)\right)}} \sim N(0,\ 1)$

Test of Significance for the difference of Standard Deviations. If s_1 and s_2 are the standard deviations of two independent samples, then under null hypothesis H_0: $\sigma_1 = \sigma_2$, *i.e.*, the sample standard deviations don't differ significantly, the statistic

$Z = \frac{s_1 - s_2}{S.E.\ (s_1 - s_2)} \sim N(0,\ 1)$ for large samples.… … … … … $(\mathbf{11.6.15})$

But in case of large samples, the S.E. of the difference of the sample standard deviations is given by

$$\text{S.E. } (s_1 - s_2) = \sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}}$$

$$\therefore\ Z = \frac{s_1 - s_2}{\sqrt{\left(\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}\right)}} \sim N(0,\ 1) \ \dots\dots\dots\dots\dots\dots . (\mathbf{11.6.16})$$

σ_1^2 and σ_2^2 are usually unknown and for large samples, we use their estimates given by the corresponding sample variances. Hence the test statistic reduces to

$$Z = \frac{s_1 - s_2}{\sqrt{\left(\frac{s_1^2}{2n_1} + \frac{s_2^2}{2n_2}\right)}} \sim N(0,\ 1) \ \dots\dots\dots\dots\dots (\mathbf{11.6.17})$$

11.7. SOLVED EXAMPLES: -

Example 11.7.1 Calculate the expectation of sample mean.

Solution: Let population has N individuals $X_1, X_2, \dots, X_N$ and sample has n units $x_1, x_2, \dots, x_n$. Then each $x_i's$ can take any of population individual values $X_1, X_2, \dots, X_N$ with equal probability 1/N. Then for every $i = 1,2, \dots, n$ we have

$$E(x_i) = X_1\frac{1}{N} + X_2\frac{1}{N} + \cdots + X_N\frac{1}{N}$$

$$= \frac{1}{N}(X_1 + X_2 + \cdots + X_N) = \frac{1}{N}(N\mu) = \mu.$$

Now$E(\bar{x}) = \frac{1}{n}\sum_{i=1}^{n} E(x_i) = \frac{1}{n}(n\,\mu) = \mu.$

Hence, sample mean is an unbiased estimate of population mean.

Example 11.7.2. Calculate the variance and standard error of sample mean.

Solution: Since each $x_i's$ can take any of population individual values $X_1, X_2, \dots, X_N$ with equal probability 1/N, therefore for every $i = 1,2, \dots, n$ we have

$$V(x_i) = E(x_i - E(x_i))^2 = E(x_i - \mu)^2$$

$$= \frac{1}{N}[(X_1 - \mu)^2 + (X_2 - \mu)^2 + \cdots + (X_N - \mu)^2] = \frac{1}{N}(N\sigma^2) = \sigma^2$$

Now, each $x_i's$ are iid r.v. and $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$, therefore

$$V(\bar{x}) = V\left(\frac{1}{n}(x_1 + x_2 + \cdots + x_n)\right) = \frac{1}{n^2}\sum_{i=1}^{n} V(x_i) = \frac{1}{n^2}(n\,\sigma^2) = \frac{\sigma^2}{n}$$

Thus,$SE(\bar{x}) = \frac{\sigma}{\sqrt{n}}$

Example 11.7.3A sample of 900 members has a mean 3.4cms. and s.d. 2.61 cms. Is the sample from a large population of mean 3.25cms, and s.d. 2.61 cms.? If the population is normal and its mean is unknown, find the 95% and 98% fiducial limits of true mean?

Solution.

Null hypothesis (H_0): The sample has been drawn from the population with mean $\mu = 3.25$ cms., ans S.D. $\sigma = 2.61$ ***cms***

Alternative Hypothesis $H_1: \mu \neq 3.25$ (Two-tailed).

Test Statistic. Under H_0, the test statistic is: $Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,\, 1)$, (since n is large).

Here, we are given $\bar{x} = 3.4cms. n = 900cms.$, $\mu = 3.25cms$ and $\sigma = 2.61cms.$

$Z = \frac{3.40-3.25}{\frac{2.61}{\sqrt{900}}} = \frac{0.15\times30}{2.61} = 1.73$.Since $|Z| < 1.96$, we conclude that the data do not provide us with any evidence against the null hypothesis (H_0) which may, therefore, be accepted at 5% level of significance.
95% Fiducial limits for the population mean μ are: $\overline{x} \pm \frac{1.96}{\sqrt{n}} \Rightarrow 3.40 \pm 1.96 \times \frac{2.61}{\sqrt{900}} \Rightarrow 3.40 \pm 0.1705, i.e.\ 3.5705\ and\ 3.2295$ 98% fiducial limits for μ are given by: $\overline{x} \pm 2.33\frac{\sigma}{\sqrt{n}}$ I.e. 3.6027 and 3.1973 .

Remark. 2.33 is the value z_1 of Z from standard normal probability integrals, such that $P(|z| > z_1) = 0.98 \Rightarrow P(Z > z_1) = 0.49$.

Example 11.7.4. An insurance agent has claimed that the average age of policyholders who insure through him is less than the average for all agents, which is 30.5 years. A random sample of 100 policyholders who had insured through him gave the following age distribution :

Age last birthday	No.of persons
12	16-20
22	21-25
20	26-30
30	31-35
16	36-40

Calculate the arithmetic mean and standard deviation of this distribution and use these values to test his claim at the 5% level of significance. You are given that $Z(1.645) = 0.95$

Solution. Null Hypothesis, $H_0: \mu = 30.5$ years, i.e., the sample mean $(\overline{x})$ and population mean (μ) do not differ significantly.
Alternative Hypothesis, H_1: μ< 30.5 years (Left-tailed alternative).

CALCULATIONS FOR SAMPLE MEAN AND S.D.

Age last birthday	No. Of persons(f)	Mid-point x	$d = \frac{x-28}{5}$	fd	fd^2
16-20	12	18	-2	-24	48
21-25	22	23	-1	-22	22
26-30	20	28	0	0	0
31-35	30	33	1	30	30
36-40	16	38	2	32	64
Total	N=100			$\sum fd$= 16	$\sum fd^2 = 164$

Thus from above table $\bar{x} = 28 + \frac{5\times16}{100} = 28.8$ years, and standard deviation of sample $s = 5 \times \sqrt{\frac{164}{100} - (\frac{16}{100} \times \frac{16}{100})} = 6.35$ years. Since the sample is large, therefore $\hat{\sigma} \simeq s = 6.35$ years. Under H_0 test statistic is $Z = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}$ = -2.681. Since $|Z| = 2.681 > 1.645$, therefore we reject the null hypothesis (Accept H_1) at 5% level of significance and conclude that the insurance agent's claim that the average age of policyholders who insure through him is less than the average for all agents, is valid.

Example 11.7.5. The means of two single large samples of 1000 and 2000 members are 67.5 inches and 68.0 inches respectively. Can the samples be regarded as drawn from the same population of standard deviation 2.5 inches? (Test at 5% level of significance).

Solution. We are given :$n_1 = 1000$, $n_2 = 2000$; $\overline{x_1} = 67.5$ inches, $\overline{x_2} = 68.0$ inches.

Null hypothesis, H_0: $\mu_1 = \mu_2$and $\sigma = 2.5\,inches$, I.e. the sample have been drawn from the same population of standard deviation 2.5 $inches$.

Alternative Hypothesis, H_1: $\mu_1 \neq \mu_2$(Two tailed).
Test Statistic. Under H_0, the test statistic is (since samples are large)

$$Z = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0,\ 1)$$

Now $$Z = \frac{67.5-68.0}{2.5\sqrt{\left(\frac{1}{1000}+\frac{1}{2000}\right)}} = \frac{-0.5}{2.5\times0.0387} = -5.1.$$

Conclusion: Since$|Z| > 3$, the value is highly significant, and we reject the null hypothesis and conclude that samples are certainly not from the same population with standard deviation 2.5.

Example 11.7.5. In a survey of buying habits, 400 women shoppers are chosen at random in super market $'A'$ located in a certain section of the city. Their average weekly food expenditure is Rs. 250 with a standard deviation of Rs. 40. For 400 women shoppers chosen at random in supermarket $'B'$ in another section of the city, the average weekly food expenditure is Rs. 220 with a standard deviation of Rs. 55. Test at 1% level of significance whether the average weekly food expenditure of the two populations of shoppers are equal.
Solution. In the usual notations, we are given that
$n_1 = 40, \overline{x_1} = Rs. 250,\ s_1 = Rs. 40, n_2 = 400, \overline{x_2} = Rs. 220,$
$s_2 = Rs. 55$

Null hypothesis, H_0: $\mu_1 = \mu_2$, $i.e.$, the average weekly food expenditures of the two populations of shoppers are equal.
Alternative hypothesis, $H_{1:}\mu_1 \neq \mu_2$, $i.e.$, (Two-tailed)

Test Statistic. Since samples are large, under H_0, the test statistic is

$$Z = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\left(\left(\frac{\sigma_1^2}{n_1}\right) + \left(\frac{\sigma_2^2}{n_2}\right)\right)}} \sim N(0,\ 1)$$

Since σ_1 and σ_2, the population standard deviations are not known, we can take for large samples: $\widehat{\sigma_1^2} = s_1^2$ and $\widehat{\sigma_2^2} = s_2^2$
And then Z is given by

$$Z = \frac{s_1 - s_2}{\sqrt{\left(\frac{s_1^2}{2n_1} + \frac{s_2^2}{2n_2}\right)}} = \frac{250 - 220}{\sqrt{\left(\frac{40^2}{400} + \frac{55^2}{400}\right)}} = 8.82(approx.)$$

Example 11.7.6. Random samples drawn from two countries gave the following data relating to the heights of adult males:

	Country *A*	**Country *B***
Mean height (in inches)	67.42	67.25
Standard deviation (in inches)	2.58	2.50
Number in samples	1000	1200

(i) Is the difference between the means significant?
(ii) Is the difference between the standard deviations significant?

Solution. We are given:

n_1 =1000, $\overline{x_1} = 67.42$ inches, $s_1 = 2.58$ inches,
$n_2 = 1200 \overline{x_2} = 67.25$inches, $s_2 = 2.50$ inches.
As in the last examples (since sample sizes are large), we can take
$\widehat{\sigma_1} = s_1 = 2.58$, $\widehat{\sigma_2} = s_2 = 2.50$.

$(\boldsymbol{i})$ H_0: $\mu_1 = \mu_2$, I.e., the sample means do not differ significantly.
H_1: $\mu_1 \neq \mu_2$ (Two tailed).
Under the null hypothesis H_0, the test statistic is $Z = \frac{\overline{x_1} = \overline{x_2}}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}} \sim$ $N(0,\ 1)$, since samples are large.
Now, $Z = \frac{67.42 - 67.25}{\sqrt{\left(\frac{(2.58)^2}{1000} + \frac{(2.50)^2}{1200}\right)}} = \frac{0.17}{\sqrt{\left(\frac{6.66}{1000} + \frac{6.25}{1200}\right)}} = 1.56$

Conclusion. Since $|Z| < 1.96$, null hypothesis may be accepted as 5% level of significance and we may conclude that there is no significant difference between sample means.

$(\boldsymbol{ii})$ Under H_0: there is no significant difference between sample standard deviations, $Z = \frac{s_1 - s_2}{S.E.\,(s_1 - s_2)} \sim N(0,\,1)$, since samples are large. Now $S.E.\,(s_1 - s_2) = \sqrt{\left(\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}\right)} = \sqrt{\left(\frac{s_1^2}{2n_1} + \frac{s_2^2}{2n_2}\right)}$. If σ_1 and σ_2 are not known and $\widehat{\sigma_1} = s_1$, $\widehat{\sigma_2} = s_2$.

$$\therefore\ S.E.(s_1 - s_2) = \sqrt{\left\{\frac{(2.58)^2}{2 \times 1000} + \frac{(2.50)^2}{2 \times 1200}\right\}} = 0.07746$$

Hence $$Z = \frac{2.58 - 2.50}{0.07746} = \frac{0.08}{0.07746} = 1.03.$$

Conclusion. Since $|Z| < 1.96$, the data don't provide us any evidence against the null hypothesis which may be accepted as 5% level of significance. Hence the sample standard deviations do not differ significantly.

Example 11.7.7. Two populations have their means equal, but S.D. of one is twice the other. Show that in the samples of size 2000 from each drawn under simple sampling conditions, the difference of means will, in all probability, not exceed 0.15σ, where σ is the smaller S.D. What is the probability that the difference will exceed half this amount?

Solution. Let the standard deviations of the two populations be σ and 2σ respectively and let μ be the mean of each of the two populations. Also, we are given $n_1 = n_2 = 2000$. If $\overline{x_1}\ and\ \overline{x_2}$ be two sample means then, since samples are large,

$$Z = \frac{(\overline{x_1} - \overline{x_2}) - E(\overline{x_1} - \overline{x_2})}{S.E.(\overline{x_1} - \overline{x_2})} \sim N(0,\,1)$$

Now $$E(\overline{x_1} - \overline{x_2}) = E(\overline{x_1}) - E(\overline{x_2}) = \mu - \mu = 0\ and$$

$$S.E.(\overline{x_1} - \overline{x_2}) = \sqrt{\left(\frac{\sigma^2}{n_1} + \frac{(2\sigma)^2}{n_2}\right)} = \sigma\sqrt{\left(\frac{1}{2000} + \frac{4}{2000}\right)} = 0.05\sigma$$

$$\therefore\ Z = \frac{(\overline{x_1} - \overline{x_2})}{S.E.(\overline{x_1} - \overline{x_2})} \sim N(0,\,1)$$

Under simple sampling conditions, we should in all probability have

$$|Z| < 3\ \Rightarrow |\overline{x_1} - \overline{x_2}| < 3\,s.e.(\overline{x_1} - \overline{x_2})$$
$$\Rightarrow |\overline{x_1} - \overline{x_2}| < 0.15\sigma,$$

which is the required result.

We want $$p = P\left[|\overline{x_1} - \overline{x_2}| > \frac{1}{2} \times 0.15\sigma\right]$$

$\therefore\ = 1 - 2P(0 \le Z \le 1.5) = 1 - 2 \times 0.4332 = 0.1336.$

11.8. SUMMARY: -

In this unit, we have studied the basic terminology of sample and population. We have also read the types of sampling. In this we have also explained about Hypothesis, Error, Level Of Significance and Procedure.

11.9.GLOSSARY:-

(i) Random Variable
(ii) Independent and Identical Random Variable
(iii) Sample
(iv) Population.
(v) Statistic.
(vi) Parameter
(vii) Estimates

11.10. REFERENCES:-

1. S. C. Gupta and V. K. Kapoor: *Fundamentals of mathematical statistics*, Sultan Chand & Sons, 2020.
2. Seymour Lipschutz and John J. Schiller *:Schaum's Outline: Introduction to Probability and Statistics*, McGraw Hill Professional, 2017.
3. J. S. Milton and J. C. Arnold: *Introduction to Probability and Statistics* (4th Edition), Tata McGraw-Hill, 2003.
4.https://www.wikipedia.org.

11.11. SUGGESTED READINGS:-

1. Rohatgi, V. K., & Saleh, A. M. E. (2015). *An introduction to probability and statistics*. John Wiley & Sons
2. A.M. Goon: *Fundamental of Statistics* (7th Edition), 1998
3. R.V. Hogg and A.T. Craig: *Introduction to Mathematical Statistics*, MacMacMillan, 2002
4. R.V. Hogg, Joseph W. Mc Kean and T. Allen: Craig: *Introduction to Mathematical Statistics* (7th edition), Pearson Education, 2013.
5. Irwin Miller and Marylees Miller John E. Freund: *Mathematical Statistics with Applications* (8th Edition). Pearson. Dorling Kindersley Pvt. Ltd. India, 2014.

11.11 TERMINAL QUESTIONS:-

TQ 1.A normal population has a mean of 0·1 and standard deviation of 2.1. Find the probability that mean of a sample of size 900 will be negative.

TQ 2. The average hourly wage of a sample of 150 workers in a plant 'A' was Rs. 2·56 with a standard deviation of Rs. 1·08. The average wage of a sample of 200 workers in plant 'B' was Rs. 2·87 with a standard deviation of Rs. 1·28. Can an applicant safely assume that the hourly wages paid by plant 'B' are higher than those paid by plant 'A' ?

TQ 3. In a certain factory there are two independent processes manufacturing the same item. The average weight in a sample of 250 items produced from one process is found to be 120 ozs. with a standard deviation of 12 ozs. while the corresponding figures in a sample of 400 items from the other process are 124 and 14. Obtain the standard error of difference between the two sample means. Is this difference significant? Also find the 99% confidence limits for the difference in the average weights of items produced by the two processes respectively.

TQ 4. The mean height of 50 male students who showed above average participation in college athletics was 68·2 inches with a standard deviation of 2·5 inches; while 50 male students who showed no interest in such participation had a mean height of 67·5 inches with a standard deviation of 2·8inches.

(i) Test the hypothesis that male students who participate in college athletics are taller than other male students.

(ii) By how much should the sample size of each of the two groups be increased in order that the observed difference of 0·7 inches in the mean heights be significant at the 5% level of significance.

11.12.ANSWER:-

Answer of Check your progress Questions:-

CYQ 1: $\binom{N}{n}$.

CYQ 2: ∞.

CYQ 3: n! ways.

CYQ 4: $1/\binom{N}{n}$.

CYQ 5: an estimator.

CYQ 6: 2.5

CYQ 7: $\frac{n}{n-1}pq$.

CYQ 8: $\sqrt{2}$.

CYQ 9: $\frac{n}{N}$.

CYQ 10: 1

CYQ 11: $\frac{\sigma^2}{n}$.

Answer of Terminal Questions:-

TQ1 0.0764
TQ2 The average hourly wages paid by plant 'B' are higher than those paid by plant 'A'.
TQ3 Standard error is 1·034. There is significant difference between the sample means. $1 \cdot 33 < |\mu_1 - \mu_2| < 6 \cdot 67$.
TQ 4 (i). College athletes are not taller than other male students.
(ii). The sample size of each of the two groups should be increased by at least 78 - 50 = 28, in order that the difference between the mean heights of the two groups is significant.

UNIT12:-EXACT SAMPLING DISTRIBUTION-I

CONTENTS:

12.1.INTRODUCTION:-

In previous unit we have explained about basics of sampling theory now in this unit we are explaining about Chi-Square Distribution and it's properties. Karl Pearson's paper of 1900 introduced what subsequently became known as the chi-squared test of goodness of fit. The terminology and allusions of 80 years ago create a barrier for the modern reader, who finds that the interpretation of Pearson's test procedure and the assessment of what he achieved are less than straightforward, notwithstanding the technical advances made since then. The psychiatrist wants to investigate whether the distribution of the patients by social class differed in these two units. She therefore erects the null hypothesis that there is no difference between the two distributions. This is what is tested by the chi squared (χ^2) test A chi-square test is used in statistics to test the independence of two events. Given the data of two variables, we can get observed count O and expected count E. Chi-Square measures how expected count E and observed count O deviates each other. The Chi-square test of independence (also known as the Pearson Chi-square test, or simply the Chi-square) is one of the most useful statistics for testing hypotheses when the variables are nominal, as often happens in clinical research.

Fig:12.1
Ref:https://en.wikipedia.org/wiki/Karl_Pearson#/media/File:Karl_Pearson,_1912.jpg
Karl Pearson
(1857-1936)

12.2.OBJECTIVES:-

After studying this unit learner will be able to:

1. Derive Chi-square distribution.
2. Explain various concepts like m.g.f., chracterstic function, etc related to chi square distribution.
3. Discuss various theorems and properties of chi-square theorem.

12.3.CHI-SQUARE(χ^2) DISTRIBUTION:-

When we consider, the null speculation is true, the sampling distribution of the test statistic is called as **Chi-squared distribution**. The Chi-squared test helps to determine whether there is a notable difference between the normal frequencies and the observed frequencies in one or more classes or categories. It gives the probability of independent variables.

As we know if X is a normal random variable with mean μ and variance σ^2, then $Z = \frac{X-\mu}{\sigma}$ is also a normal variable with mean 0 and variance 1. Here the Z variable is known as standard normal variate. The square of Z is called a chi-square variable with 1 degree of freedom i.e. if $X \sim N(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$.
Consequently, $Z^2 = \left(\frac{X-\mu}{\sigma}\right)^2$, is a chi-square variate with degree of freedom 1. In place of Z^2 we use the symbol χ^2.

- A teacher-researcher has obtained data from a questionnaire used for teachers and head teachers in the form of a contingency table of 3 × 3 for anxiety and awareness levels. Chi-Square(χ^2)test will indicate whether the two variables are interdependent.

In general, if X_i, $(i = 1,2, \dots, n)$ are n independent normal variables with mean μ_i and variance σ_i^2, $(i = 1,2, \dots, n)$, then

$$\chi^2 = \sum_{i=1}^{n} \left(\frac{X_i - \mu_i}{\sigma_i}\right)^2, \text{ ....(13.3.1)}$$

is a chi-square variable with n degree of freedom.

Problem 12.3.1: Let χ^2 be a chi-square variable with n degree of freedom. Obtain the distribution of χ^2.

Solution: Let $X_i, (i = 1,2, \dots, n)$ are independent normal distributions $N(\mu_i, \sigma_i^2)$, we want the distribution of

$$\chi^2 = \sum_{i=1}^{n} \left(\frac{X_i - \mu_i}{\sigma_i}\right)^2 = \sum_{i=1}^{n} U_i^2, \text{ where } U_i = \frac{X_i - \mu_i}{\sigma_i}$$

Since $X_i's$ are independent, $U_i's$ are also independent and follow $N(0,1)$.

$$M_{\chi^2}(t) = M_{\sum U_i^2}(t) \quad = \prod_{i=1}^{n} M_{U_i^2}(t) = \left[M_{U_i^2}(t)\right]^n.$$

Now, $M_{U_i^2}(t) = E[\exp[tU_i^2]] = \int_{-\infty}^{\infty} \exp(tu_i^2)\, f(x_i) dx_i$

$$= \int_{-\infty}^{\infty} \exp(tu_i^2) \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right) dx_i$$

$$= \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} \exp(tu_i^2) \exp\left(-\frac{u_i^2}{2}\right) du_i \because \left[u_i = \frac{x_i - \mu}{\sigma}\right]$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left\{-\left(\frac{1-2t}{2}\right)\mu_i^2\right\} d\mu_i$$

$$= \frac{1}{\sqrt{2\pi}} \frac{\sqrt{\pi}}{\left(\frac{1-2t}{2}\right)^{\frac{1}{2}}} = (1-2t)^{-\frac{1}{2}}$$

$$\therefore\ M_{\chi^2}(t) = (1-2t)^{-\frac{n}{2}}$$

The above equation is moment generating function of a Gamma variable with parameters $\frac{1}{2}$ and $\frac{1}{2}n$. Therefore from the uniqueness of moment generating function, χ^2 is a Gamma variate with parameter $\frac{1}{2}$ and $\frac{1}{2}n$. Hence if a random variable Y has a chi-square distribution with n degree of freedom, then we write $Y \sim \chi^2(n)$ and its p.d.f, is given by :

$$f(y) = \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} e^{-\frac{y}{2}} y^{\left(\frac{n}{2}\right)-1}, \qquad 0 \leq y < \infty.$$

Example 12.3.2: If $Y \sim \chi^2(n)$, then $\frac{Y}{2} \sim \gamma\left(\frac{n}{2}\right)$.

Proof. Let $W = \frac{1}{2}Y$. Then $Y = 2W$.

$$g(w) = f(y)\left|\frac{dy}{dw}\right|$$

$$= \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} e^{-\frac{y}{2}} \cdot (y)^{\frac{n}{2}-1} \cdot 2$$

$$= \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} e^{-w} \cdot (2w)^{\frac{n}{2}-1} \cdot 2$$

$$= \frac{1}{\Gamma\left(\frac{n}{2}\right)} e^{-w} \cdot w^{\frac{n}{2}-1};\ 0 \leq y < \infty$$

Therefore, $W \sim \gamma\left(\frac{n}{2}\right)$.

12.4.M.G.F OF CHI-SQUARE (χ^2) DISTRIBUTION: -

Let $X \sim \chi^2(n)$. Then M.G.F. of X is:

$$M_X(t) = E(e^{tX}) = \int_0^\infty e^{tx} f(x)dx$$

$$= \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} \int_0^\infty e^{tx} e^{-\frac{x}{2}} (x)^{\frac{n}{2}-1} dx$$

$$= \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} \int_0^\infty exp\left[-\left(\frac{1-2t}{2}\right)x\right](x)^{\frac{n}{2}-1} dx$$

$$= \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} \times \frac{\Gamma\left(\frac{n}{2}\right)}{\left(\frac{1-2t}{2}\right)^{\frac{n}{2}}}$$

Hence, the M.G.F. of $\boldsymbol{\chi^2}$**-** variate with n degree of freedom is:

$$M_X(t) = \left(\frac{1-2t}{2}\right)^{-\frac{n}{2}}, |2t| < 1.$$

- Cumulant Generating Function of $\chi^2 -$ distribution$K_X(t) = logM_X(t)$.
- $k_r =$ Coefficient of $\frac{t^r}{r!}$ in $K(t) = n2^{r-1}(r-1)!$.
- Mode of the chi-square distribution with$n$ degree of freedom is $(n-2)$.
- Mean of the chi-square distribution$k_1 = n$, Variance $\mu_2 = k_2 = 2n$.
- $\mu_3 = k_3 = 8n$, $\mu_4 = k_4 + 3k_2^2 = 48n + 12n^2$.

- $\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{8}{n}$ and $\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{12}{n} + 3$,
- Skewness = $\sqrt{\frac{2}{n}}$.
- Chi square distribution tends to normal distribution for large $d.f.$
- The sum of independent Chi – square variates is also a χ^2 −variate. Converse is also true.

12.5. THEOREMS ONCHI-SQUARE (χ^2) DISTRIBUTION:-

Theorem 12.5.1: If $\mathcal{X}_1^2$ and $\mathcal{X}_2^2$ are two independents $\mathcal{X}^2$-variates with n_1 and n_2d.f. respectively, then $\frac{\mathcal{X}_1^2}{\mathcal{X}_2^2}$ is a $\beta_2\left(\frac{n_1}{2},\frac{n_2}{2}\right)$ variate.

Proof: Since $\mathcal{X}_1^2$ and $\mathcal{X}_2^2$ are independent $\mathcal{X}^2$-variates with n_1 and n_2 d.f.
respectively, therefore their joint probability differential is given by the compound probability theorem as

$$dP(\mathcal{X}_1^2,\mathcal{X}_2^2) = dP(\mathcal{X}_1^2)dP(\mathcal{X}_2^2)$$

$$= \frac{1}{2^{\frac{n_1}{2}}\Gamma\left(\frac{n_1}{2}\right)} e^{-\frac{\mathcal{X}_1^2}{2}}(\mathcal{X}_1^2)^{\left(\frac{n_1}{2}\right)-1}d(\mathcal{X}_1^2).\frac{1}{2^{\frac{n_2}{2}}\Gamma\left(\frac{n_2}{2}\right)} e^{-\frac{\mathcal{X}_2^2}{2}}(\mathcal{X}_2^2)^{\left(\frac{n_2}{2}\right)-1}d(\mathcal{X}_2^2)$$

$$= \frac{1}{2^{\frac{(n_1+n_2)}{2}}\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)} e^{-\frac{(\mathcal{X}_1^2+\mathcal{X}_2^2)}{2}}(\mathcal{X}_1^2)^{\left(\frac{n_1}{2}\right)-1}(\mathcal{X}_2^2)^{\left(\frac{n_2}{2}\right)-1}d(\mathcal{X}_1^2)d(\mathcal{X}_2^2)$$

Now, let $u = \frac{\mathcal{X}_1^2}{\mathcal{X}_2^2}$ and $v = \mathcal{X}_2^2$. Then $\mathcal{X}_1^2 = uv$ and $\mathcal{X}_2^2 = v$. Thus, the Jacobian is given by

$$J = \frac{\partial(\mathcal{X}_1^2,\mathcal{X}_2^2)}{\partial(u,v)} = \begin{vmatrix} v & u \\ 0 & 1 \end{vmatrix} = v.$$

Thus, the joint distribution of random variables U and V becomes

$dG(u,v)$

$$= \frac{1}{2^{\frac{(n_1+n_2)}{2}}\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)} e^{\frac{\left(-\frac{(1+u)v}{2}\right)}{2}}(uv)^{\left(\frac{n_1}{2}\right)-1}(v)^{\left(\frac{n_2}{2}\right)-1}|J|d(u)d(v)$$

$$= \frac{1}{2^{\frac{(n_1+n_2)}{2}}\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)} e^{\frac{\left(-\frac{(1+u)v}{2}\right)}{2}}(u)^{\left(\frac{n_1}{2}-1\right)}(v)^{\left(\frac{n_1+n_2}{2}-1\right)}d(u)d(v)$$

Integrating w.r.t. v over the range 0 to ∞, we get the marginal distribution of U as

$$d\,G_1(u) = \int_0^\infty dG(u,v)$$

$$= \frac{1}{2^{\frac{(n_1+n_2)}{2}}\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)}(u)^{\left(\frac{n_1}{2}-1\right)}du \times \int_0^\infty e^{\frac{\left(-\frac{(1+u)v}{2}\right)}{2}}(v)^{\left(\frac{n_1+n_2}{2}-1\right)}dv$$

$$= \frac{1}{2^{\frac{(n_1+n_2)}{2}}\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)}(u)^{\left(\frac{n_1}{2}-1\right)}du \times \int_0^\infty e^{\frac{\left(-\frac{(1+u)v}{2}\right)}{2}}(v)^{\left(\frac{n_1+n_2}{2}-1\right)}dv$$

$$= \frac{u^{(n_1/2)-1}}{2^{\frac{(n_1+n_2)}{2}}\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)}du \times \frac{\Gamma\left(\frac{n_1+n_2}{2}\right)}{(1+u)^{(n_1+n_2)/2}}$$

$$= \frac{1}{B\left(\frac{n_1}{2},\frac{n_2}{2}\right)} \times \frac{u^{(n_1/2)-1}}{(1+u)^{(n_1+n_2)/2}}du, \quad 0 \le u \le \infty.$$

Hence $U = \frac{\mathcal{X}_1^2}{\mathcal{X}_2^2}$ is a $\beta_2\left(\frac{n_1}{2},\frac{n_2}{2}\right)$ variate.

Theorem 12.5.2: If $\mathcal{X}_1^2$ and $\mathcal{X}_2^2$ are two independents $\mathcal{X}^2$-variates with n_1 and n_2d.f. respectively, then

$$U = \frac{\mathcal{X}_1^2}{\mathcal{X}_1^2 + \mathcal{X}_2^2} \text{ and } V = \mathcal{X}_1^2 + \mathcal{X}_2^2$$

Are independently distributed, U as $\beta_1\left(\frac{n_1}{2},\frac{n_2}{2}\right)$ variate and V as $\mathcal{X}^2$-variates with (n_1+n_2)d.f.

Proof: As we discussed in previous theorem

$$dP(\mathcal{X}_1^2,\mathcal{X}_2^2) = \frac{1}{2^{\frac{(n_1+n_2)}{2}}\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)}e^{-\frac{(\mathcal{X}_1^2+\mathcal{X}_2^2)}{2}}(\mathcal{X}_1^2)^{\left(\frac{n_1}{2}\right)-1}(\mathcal{X}_2^2)^{\left(\frac{n_2}{2}\right)-1}d(\mathcal{X}_1^2)d(\mathcal{X}_2^2),$$

$0 \le \mathcal{X}_1^2, \mathcal{X}_2^2 < \infty$

Now, let $u = \frac{\mathcal{X}_1^2}{\mathcal{X}_1^2+\mathcal{X}_2^2}$ and $v = \mathcal{X}_1^2 + \mathcal{X}_2^2$. Then $\mathcal{X}_1^2 = uv$ and $\mathcal{X}_2^2 = v - \mathcal{X}_1^2 = (1-u)v$.

As $\mathcal{X}_1^2$ and $\mathcal{X}_2^2$both range from 0 to ∞, therefore u ranges from 0 to 1 and v from 0 to ∞. And the Jacobian is given by

$$J = \frac{\partial(\mathcal{X}_1^2,\mathcal{X}_2^2)}{\partial(u,v)} = \begin{vmatrix} v & u \\ -v & 1-u \end{vmatrix} = v.$$

Thus, the joint distribution of random variables U and V becomes

$$dG(u,v) = \frac{1}{2^{\frac{(n_1+n_2)}{2}}\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)} e^{\frac{-v}{2}}(uv)^{\left(\frac{n_1}{2}\right)-1}((1 - u)v)^{\left(\frac{n_2}{2}\right)-1}|J|d(u)d(v)$$

$$= \frac{1}{2^{\frac{(n_1+n_2)}{2}}\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)}(u)^{\left(\frac{n_1}{2}-1\right)}(1 - u)^{\left(\frac{n_2}{2}\right)-1}e^{\frac{-v}{2}}(v)^{\left(\frac{n_1+n_2}{2}-1\right)}dudv$$

$$= \left[\frac{\Gamma\left(\frac{n_1+n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)}(u)^{\left(\frac{n_1}{2}-1\right)}(1-u)^{\left(\frac{n_2}{2}\right)-1}du\right] \times \frac{1}{2^{\frac{(n_1+n_2)}{2}}\Gamma\left(\frac{n_1+n_2}{2}\right)} e^{\frac{-v}{2}}(v)^{\left(\frac{n_1+n_2}{2}-1\right)}dv$$

Since the joint probability differential of U and V is the product of their respective probabitlity differentials, U and V are independently distributed, with

$$d\,G_1(u) = \frac{1}{B\left(\frac{n_1}{2},\frac{n_2}{2}\right)}u^{(n_1/2)-1}(1-u)^{(n_2/2)-1}du,\ \ 0 \le u \le 1$$

$$d\,G_2(v) = \frac{1}{2^{(n_1+n_2)/2}.\Gamma\left(\frac{n_1+n_2}{2}\right)}\exp\left(-\frac{v}{2}\right)v^{\left[\frac{n_1+n_2}{2}\right]-1}\ dv,\quad 0 \le v \le \infty$$

$i.e.,$ U as a $\beta_1\left(\frac{n_1}{2},\frac{n_2}{2}\right)$ variate and V as a $\mathcal{X}^2$–variate with (n_1+n_2) d.f.

Remark 12.5.3 On the basis of above we have following remarks:
If $X \sim \mathcal{X}^2{}_{(n_1)}$ and $Y \sim \mathcal{X}^2{}_{(n_2)}$ are independent chi-square variates then:

(i) $X + Y \sim \mathcal{X}^2{}_{(n_1+n_2)}$, hence the sum of two independent chi-square variates is also a chi-square variate.

(ii) $\frac{X}{Y} \sim \beta_2\left(\frac{n_1}{2},\frac{n_2}{2}\right)$, hence the ratio of two independent chi-square variates is a β_2 variate.

Theorem 12.5.4 In a random and large sample,

$$\mathcal{X}^2 = \sum_{i=1}^{k}\left[\frac{(n_i - np_i)^2}{np_i}\right],$$

Follows chi-square distribution approximately with $(k-1)$ degrees of freedom, where n_i is the observed frequency and np_i the corresponding expected frequency of the ith class, $(i = 1,2,\ldots.k), \sum_{i=1}^{k} n_i = n$.

Proof. Let us consider a random sample of size n, whose members are distributed at random in k classes or cells. Let p_i be the probability that sample observation will fall in the ith cell, $(i = 1,2,\ldots.k)$. Then the probability P of there being n_i members in the ith cell, $(i =$

$1,2, \ldots . k)$ respectively is given by the multinomial probability law, by the expression

$$P = \frac{n\,!}{n_1\,!\ n_2\,! \ldots .\ n_k\,!} p_1^{\,n_1} p_2^{\,n_2} \ldots p_k^{\,n_k},$$

where $\sum_{i=1}^{k} n_i = n$ and $\sum_{i=1}^{k} p_i = 1$.

If n is sufficiently large so that $n_i, (i = 1,2, \ldots .., k)$ are not small then by Stirling's approximation to factorials for large n, which is

$\lim_{n\to\infty}(n\,!) \approx \sqrt{2\pi}\, e^{-n}\, n^{n+\frac{1}{2}}$, we get

$$P \approx \frac{\sqrt{2\pi}\, e^{-n}\, n^{n+\frac{1}{2}}}{(\sqrt{2\pi})^k e^{-(n_1+n_2+\cdots\ldots+n_k)}} \times \frac{p_1^{\,n_1} p_2^{\,n_2} \ldots p_k^{\,n_k}}{n_1^{\,n_1+\frac{1}{2}}\ n_2^{\,n_2+\frac{1}{2}} \ldots .\ n_k^{\,n_k+\frac{1}{2}}}$$

$$\approx \frac{e^{-1}\, n^{n+\frac{1}{2}} \left(\frac{np_1}{n_1}\right)^{n_1+\frac{1}{2}} \left(\frac{np_2}{n_2}\right)^{n_2+\frac{1}{2}} \ldots \left(\frac{np_k}{n_k}\right)^{n_k+\frac{1}{2}}}{\left(\sqrt{2\pi}\right)^{k-1} e^{-n} n^{n_1+n_2+\cdots\ldots+n_k+\left(\frac{k}{2}\right)} (p_1 p_2 \ldots . p_k)^{\frac{1}{2}}}$$

$$\approx C \prod_{i=1}^{k} \left(\frac{np_i}{n_i}\right)^{n_1+\frac{1}{2}}$$

where $C = \frac{1}{(2\pi)^{(k-1)/2} n^{(k-1)/2} (p_1 p_2 \ldots p_k)^{1/2}}$, is a constant, independent of n_i's.

therefore,

$\log P \approx \log C + \sum_{i=1}^{k} (n_i + \frac{1}{2}) \log\left(\frac{\lambda_i}{n_i}\right)$,

$$\log P/C \approx \sum_{i=1}^{k} (n_i + \frac{1}{2}) \log\left(\frac{\lambda_i}{n_i}\right),$$

where $\lambda_i = nP_i$ is the expected frequency for the ith cell, $i.e.$,

$$E(n_i) = nP_i = \lambda_i,\ (i = 1,2, \ldots k).$$

Define $\xi_i = \frac{n_i - \lambda_i}{\sqrt{\lambda_i}}$,

so that $n_i - \lambda_i = \xi_i \sqrt{\lambda_i} \quad \Rightarrow \quad n_i = \lambda_i + \xi_i \sqrt{\lambda_i}$.

This implies that

$$\log(\text{P/C}) \approx \sum_{i=1}^{k} \left(\lambda_i + \xi_i \sqrt{\lambda_i} + \tfrac{1}{2}\right) \log\left[\frac{\lambda_i}{\lambda_i + \xi_i \sqrt{\lambda_i}}\right]$$

$$\approx \sum_{i=1}^{k} (\lambda_i + \xi_i \sqrt{\lambda_i} + \frac{1}{2}) \log[1/\{1 + \xi_i/\sqrt{\lambda_i}\}]$$

$$= -\sum_{i=1}^{k} (\lambda_i + \xi_i \sqrt{\lambda_i} + \frac{1}{2}) \log[\{1 + (\xi_i/\sqrt{\lambda_i})\}]$$

Further, if we assume that ξ_i is small compared with λ_i, the expansion of $\log 1 + (\xi_i/\sqrt{\lambda_i})$ in ascending powers of $\xi_i/\sqrt{\lambda_i}$ is valid.

$$\therefore \quad \log(\text{P/C}) \approx -\sum_{i=1}^{k} \left(\lambda_i + \xi_i \sqrt{\lambda_i} + \tfrac{1}{2}\lambda_i\right) \left[\frac{\xi_i}{\sqrt{\lambda_i}} - \frac{1}{2}\frac{\xi_i^{\,2}}{\lambda_i} + O\,(1/\lambda_i^{\,3/2})\right]$$

$$\approx -\sum_{i=1}^{k}\left[\,\xi_i\sqrt{\lambda_i}-\frac{1}{2}\,\xi_i^{\,2}+O\,(1/\lambda_i^{\,-1/2})\right],$$

Neglecting higher powers of $\xi_i\sqrt{\lambda_i}$ if ξ_i is small compared with λ_i.

Since n is large, so is $\lambda_i = nP_i$. Hence $O\left(\lambda_i^{-\frac{1}{2}}\right) \to 0$ for large n.

Also $\sum_{i=1}^{k} \xi_i\sqrt{\lambda_i} \approx \sum_{i=1}^{k}(n_i-\lambda_i) = \sum_{i=1}^{k} n_i - \sum_{i=1}^{k}\lambda_i$

$\approx \sum_{i=1}^{k} n_i - n\sum_{i=1}^{k} P_i = n - n = 0$ $(\because \sum n_i = n, \sum P_i = 1)$

$\therefore$Log(P/C) $\approx -\left[\sum_{i=1}^{k} \xi_i\sqrt{\lambda_i} + \frac{1}{2}\sum_{i=1}^{k} \xi_i^{\,2} + O\left(\frac{1}{\lambda_i^{-\frac{1}{2}}}\right)\right] \approx -\frac{1}{2}\sum_{i=1}^{k} \xi_i^{\,2}$

$\Rightarrow$P$\approx C$ exp $\left(-\frac{1}{2}\sum_{i=1}^{k} \xi_i^{\,2}\right)$

Which shows that ξ_i, $(i = 1,2, \ldots k)$ are distributed as independent standard normal varieties.

Hence $\sum_{i=1}^{k} \xi_i^{\,2} = \sum_{i=1}^{k}\left[\frac{(n_i-\lambda_i)^2}{\lambda_i}\right]$,

Being the sum of the squares of k independent standard normal variates is a $\mathcal{X}^2$variate with $(k-1)$ d.f., one d.f. being lost because of the linear constraint

$$\sum_{i=1}^{k} \xi_i\sqrt{\lambda_i} = \sum(n_i-\lambda_i) = 0 \;\Rightarrow \sum_{i=1}^{k} n_i = \sum_{i=1}^{k}\lambda_i$$

Remarks 12.5.5(i). If O_i and $E_i(i = 1,2, \ldots, k)$, be a set of observed and expected frequencies, then

$$\mathcal{X}^2 = \sum_{i=1}^{k}\left[\frac{(O_i-E_i)^2}{E_i}\right], \qquad (\sum_{i=1}^{k} O_i = \sum_{i=1}^{k} E_i)$$

Follows chi-square distribution with $(k-1)$ degree of freedom.
Another convenient form of this formula is as follows:

$$\mathcal{X}^2 = \sum_{i=1}^{k}\left[\frac{O_i^{\,2}-E_i^{\,2}-2O_iE_i}{E_i}\right] = \sum_{i=1}^{k}\left(\frac{O_i^{\,2}}{E_i}+E_i-2O_i\right)$$

$$= \sum_{i=1}^{k}\left(\frac{O_i^{\,2}}{E_i}\right) + \sum_{i=1}^{k} E_i - 2\sum_{i=1}^{k} O_i$$

$$\approx \sum_{i=1}^{k}\left(\frac{O_i^{\,2}}{E_i}\right) - N,$$

Where$\sum_{i=1}^{k} O_i = \sum_{i=1}^{k} E_i = N$ (say), is the total frequency.

(ii). Conditions for the Validity of $\mathcal{X}^2$-test: $\mathcal{X}^2$-test is an approximate test for large values of n. For the validity of chi-square test of '*goodness of fit*' between theory and experiment, the following conditions must be satisfied:

a) The sample observations should be independent.

b) Constraints on the cell frequencies, if any, should be linear, $e.g., \sum n_i = \sum \lambda_i$ or$\sum O_i = \sum E_i$.

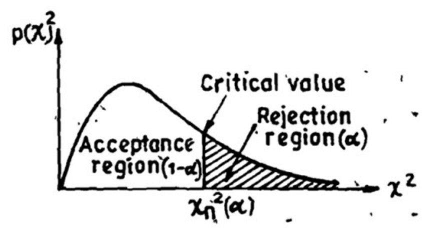
c) N, the total frequency should be reasonably large, say, greater than 50.

d) No theoretical cell frequency should be less than 5. (The chi square distribution is essentially a continuous distribution but it cannot maintain its character of continuity if cell frequency is less than 5). If any theoretical cell frequency is less than 5, then for the application of $\mathcal{X}^2$-test, it is pooled with the preceding or succeeding frequency so that the pooled-frequency is more than 5 and finally adjust for the degree of freedom lost in pooling.

e) It may be noted that the $\mathcal{X}^2$-test depends only on the set of observed and expected frequencies and on degrees of freedom $(d.f.)$. It does not make any assumptions regarding the parent population from which the observations are taken. Since $\mathcal{X}^2$defined in (13.8) does not involve any population parameters, it is termed as a statistic and the test is known as *Non-Parametric Test* or *Distribution-Free Test*.

f) **Critical Values.** Let ${\mathcal{X}_n}^2(\alpha)$denote the value of chi-square for n.$d.f$. such that the area to the right of this point is α, i.e.,

$$P[\mathcal{X}^2 > {\mathcal{X}_n}^2(\alpha)] = \alpha$$



The value ${\mathcal{X}_\text{n}}^2(\alpha)$defined in previous equation is known as the upper (right − tailed) α-point or Critical Value or Significant Value of chi-square for n degree of freedom and has been tabulated for different values of n and α in Table VI in the Appendix at the end of the book. From these tables we observe that the critical values of $\mathcal{X}^2$ increase as n (d.f.) increases and level of significance (α) decreases.

12.6. APPLICATION OF CHI-SQUARE (TEST IN CHI-SQUARE):-

Chi-square Test for Population Variance. Suppose we want to test if a random sample $x_i, (i = 1,2, \dots, n)$ has been drawn from a normal population with a specified variance $\sigma^2 = \sigma_0^2$, (say).

Under the null hypothesis that the population variance is $\sigma^2 = \sigma_0^2$, the statistic

$$\chi^2 = \sum_{i=1}^{n}\left[\frac{(x_i-\overline{x})^2}{\sigma_0^2}\right] = \frac{1}{\sigma_0^2}\left[\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}\right] = \frac{ns^2}{\sigma_0^2}$$

Follows chi-square distribution with $(n-1)d.f.$

By comparing the calculated value with the tabulated value of χ^2 for $(n-1)\ d.f.$ at certain level of significance, (usually 5%), we may retain or reject the null hypothesis.

Remarks 12.6.1 (i). The above test can be applied only if the population from which sample is drawn is normal.

(ii). If the sample size n is large (> 30), then we can use Fisher's approximation

$$\sqrt{2\chi^2} \sim N\left(\sqrt{2n-1},\ 1\right)$$

i.e., $$Z = \sqrt{2\chi^2} - \sqrt{2n-1} \sim N(0,1)$$

And apply Normal Test.

Note on Degrees of Freedom $(\mathbf{d.f.})$ The number of independent variates which make up the statistic (e.g.,χ^2) is known as the degrees of freedom (d.f.) and is usually denoted by ν. The number of degrees of freedom, in general, is the total number of observations less the number of independent constraints imposed on the observations. For example, if k is the number of independent constraints in a set of data of n observations then $\nu = (n-k)$. Thus, in a set of n observations usually, the degrees of freedom for χ^2 are $(n-1)$, one d.f. being lost because of the linear constraint $\sum_i O_i = \sum_i E_i = N$, on the frequencies. If 'r' independent linear constraints are imposed on the cell frequencies, then the d.f. are reduced by 'r'. In addition, if any of the population parameter(s) is(are) calculated from the given data and used for computing the expected frequencies then in applying $\chi^2 - test$ of goodness of fit, we have to subtract one d.f. for each parameter calculated. Thus if 's' is the number of population parameters estimated from the sample observations (n in number), then the required number of degrees of freedom for $\chi^2 - test$ is $(n-s-1)$.If any one or more of the theoretical frequencies is less than 5 then in applying, $\chi^2 - test$ we have also to subtract the degrees of freedom lost in pooling these frequencies with the preceding or succeeding frequency (or frequencies). In a $r \times s$ cotingency table, in calculating the expected frequencies, the row totals, the column totals and the grand totals remain fixed. The fixation of $'r'$ column totals and $'s'$ row totals impose $(r+s)$ constraints on the cell frequencies. But since $\sum_{i=1}^{r} A_i = \sum_{i=1}^{r}(B_j) = N$. The total number of independent constraints is only $(r+s-1)$. Further, since the total number of the cell-frequencies is $r \times s$, the required number of degrees of freedom is $\nu = rs - (r+s-1) = (r-1)(s-1)$.

Chi − square Test of Goodness of Fit. A very powerful test for testing the significance of the discrepancy between theory and experiment was given by Prof. Karl Pearson in 1900 and is known as "Chi-square Test of Goodness of fit." It enables us to find if the deviation of the experiment from theory is just by chance or is it really due to the inadequacy of the theory to fit the observed data.If O_i, $(i = 1,2,\dots,n)$ is a set of observed (experimental) frequencies and $E_i(i = 1,2,3,\dots,n)$ is the corresponding set of expected (theoretical or hypothetical)

Frequencies, then Karl Pearson's chi-square, given by

$$\chi^2 = \sum_{i=0}^{n}\left[\frac{(O_i - E_i)^2}{E_i}\right], \qquad \left(\sum_{i=0}^{n} O_i = \sum_{i=0}^{n} E_i\right)$$

Follows chi-square distribution with $(n-1)d.f.$

Chi-square Probability Curve

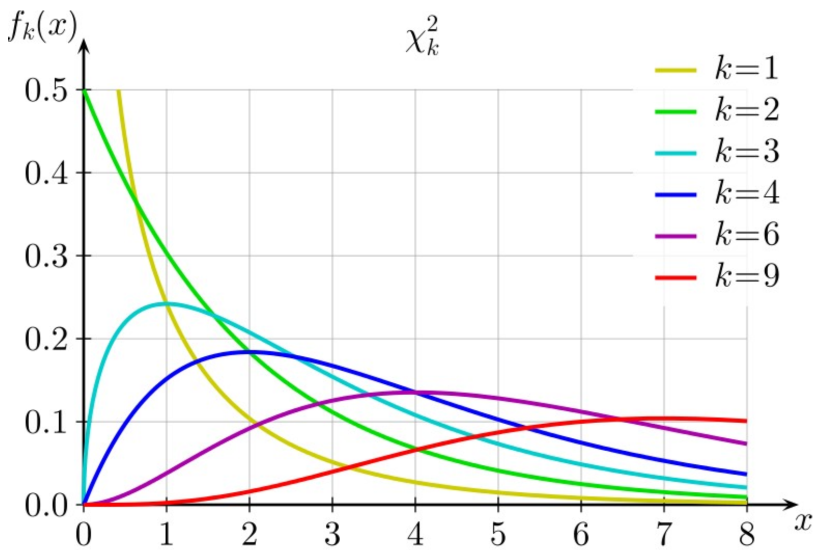


Fig 12.6.1

Ref: https://en.wikipedia.org/wiki/Chi-squared_distribution#/media/File:Chi-square_pdf.svg

Check Your Progress

1) In Chi-Square test the sample observations should be…………..
2) No theoretical cell frequency should be less than……….
3) A researcher asked 933 people what their favourite type of TV programme was: news, documentary, soap or sports. They could only choose one answer. As such, the researcher had the number of people who chose each category of programme. How should she analyse these data? (i) t-test (ii) One-way analysis of variance (iii) Chi-square test (iv) Regression

4) Chi-square is used to analyse:

(i) Scores (ii)Ranks (iii) Frequencies (iv)Any of these

5) On which of the following does the critical value for a chi-square statistic rely?

(i)The degrees of freedom (ii)The sum of the frequencies (iii) The row totals(iv) The number of variables

12.7.SOLVED EXAMPLES: -

Example12.7.1: It is believed that the precision (as measured by variance) of an instrument is no more than 0.16. Write down the null and alternative hypothesis for testing this belief. Carry out the test at 1% level, given 11 measurements of the same subject on the instrument:

2.5, 2.3, 2.4, 2.3, 2.5, 2.7, 2.5, 2.6, 2.7, 2.5.

Solution. Null Hypothesis, $H_0 = \sigma^2 = 0.16$
Alternative hypothesis: $H_1: \sigma^2 > 0.16$

COMPUTAION OF SAMPLE VARIANCE

X	$X - \overline{X}$	$(X - \overline{X})^2$
2.5	-0.01	0.0001
2.3	-0.21	0.0441
2.4	-0.11	0.0121
2.3	-0.21	0.0441

2.5	-0.01	0.0001
2.7	0.19	0.0361
2.5	-0.01	0.0001
2.6	0.09	0.0081
2.6	0.09	0.0081
2.7	0.19	0.0361
2.5	-0.01	0.0001
$\overline{X} = \frac{27.6}{11} = 2.51$		$\sum(X-\overline{X})^2 = 0.1891$

Under the null hypothesis H_0: $\sigma^2 = 0.16$, the test statistic is:

$$\chi^2 = \frac{ns^2}{\sigma^2} = \frac{\sum(X-\overline{X})^2}{\sigma^2} = \frac{0.1891}{0.16} = 1.182$$

Which follows χ^2-distribution with d.f. $(11-1) = 10$.

Since the calculated value of χ^2 is less than the tabulated value $23.2\ of\ \chi^2$ for 10 $d.f.$ at 1% level of significance, it is not significant. Hence H_0 may be accepted and we conclude that the data are consistent with the hypothesis that the precision of the instrument is 0.16 .

Example 12.7.2 To sample polls of votes for two candidates $A\ and\ B$ for a public office are taken, one from among the residents of rural areas. The results are given in the table. Examine whether the nature of the area is related to voting preference in this election.

Votes for Area	**A**	**B**	**Total**
Rural	620	380	1000
Urban	550	450	1000
Total	1170	830	2000

Solution. In a 2 × 2 contingency table, $d.f. = (2-1)(2-1) = 1$,Under the null hypothesis that the nature of the area is independent of the voting preference in the election, we get the observed frequencies as follows:

$$E(620) = \frac{(1170 \times 1000)}{2000} = 585, \qquad E(380) = \frac{(830 \times 1000)}{2000} = 415,$$

$$E(550) = \frac{(1170 \times 1000)}{2000} = 585,\ and\ E(450) = \frac{(830 \times 1000)}{2000} = 415,$$

$$\therefore \chi^2 = \sum\left[\frac{(O-E)^2}{E}\right]$$

$$= \frac{(620-585)^2}{585} + \frac{(380-415)^2}{415} + \frac{(550-585)^2}{585} + \frac{(450-415)^2}{415}$$

$$= 35^2\left[\frac{1}{585} + \frac{1}{415} + \frac{1}{585} + \frac{1}{415}\right]$$

$$= 1125[2 \times 0.002409 + 2 \times 0.001709] = 10.0891$$

Tabulated $\chi^2_{0.05}\, for\ (2-1)(2-1) = 1\ d.f.\ is$ 3.841. Since calculated χ^2 is much greater than the tabulated value, it is highly significant and null hypothesis is rejected at 5% level of significance. Thus, we conclude that nature of area is related to voting preference in the election.

Example 12.7.3 Test the hypothesis that $\sigma = 10$, given that $s = 15$ for a random sample of size 50 from a normal population.

***Solution*.** Null Hypothesis, $H_0: \sigma = 10$.

We are given $n = 50$, $s = 15$ $\therefore$ $\chi^2 = \frac{ns^2}{\sigma^2} = \frac{(50\times225)}{100} = 112.5$

Since n is large, using $(13.14a)$, the test statistic is $Z = \sqrt{2\chi^2} - \sqrt{2n-1} \sim N(0,\ 1)$.Now, $Z = \sqrt{225} - \sqrt{99} = 15 - 9.95 = 5.05$.Since $|Z| > 3$, it is significant at all levels of significance and hence H_0 is rejected and we conclude that $\sigma \neq 10$.

Example 12.7.4 The following figures show the distribution of digits in numbers chosen at random from a telephone directory. Test whether the digits may be taken to occur equally frequently

Digits:	0	1	2	3	4	5	6	7	8	9	Total
Frequency:	1026	1107	997	966	1075	933	1107	972	964	853	10,000

Solution. Here we set up the null hypothesis that the digits occur equally frequently in directory.

Under the null hypothesis, the expected frequency for each of the digits $0,\ 1,\ 2, \ldots, 9\ is\ \frac{1000}{10} = 100$. The value of χ^2 is computed as follows:

$$CALCULATIONS\ FOR\ \chi^2$$

Digits	Observed Frequency (O)	Expected Frequency (E)	$(O-E)^2$	$\frac{(O-E)^2}{E}$
0	1026	1000	676	0.676
1	1107	1000	11449	11.449
2	997	1000	9	0.009
3	966	1000	1156	1.156
4	1075	1000	5625	5.625
5	933	1000	4489	4.489
6	1107	1000	11449	11.449
7	972	1000	784	0.784
8	964	1000	1296	1.296
9	853	1000	21609	21.609
$Total$	10,000	10,000		58.542

$$\therefore\ \chi^2 = \sum\left[\frac{(O-E)^2}{E}\right] = 58.542$$

The number of degrees of freedom $= 10 - 1 = 9$, (since we are given 10 frequencies subjected to only one linear constraint ($\sum O = \sum E = 10{,}000$).

The tabulated $\chi^2_{0.05}$ for 9 $d.f.$ =16.919

Since the calculated χ^2 is much greater than the tabulated value, it is highly significant and we reject the null hypothesis. Thus, we conclude that the digits are not uniformly distributed in the directory.

Example 12.7.5. The following table gives the number of aircraft accidents that occur during the various days of the week.

Days:	Sun.	Mon.	Tues.	Wed.	Thus.	Fri.	Sat.
No. of Accidents:	14	16	8	12	11	9	14

(Given: The values of chi-square significant at $5, 6, 7,\ d.f.$ are respectively'$11.07,\ 12.59,\ 14.07$ at the 5% level of significance.)

Solution. Here we setup the null hypothesis that the accidents are uniformly distributed over the week.Under the null hypothesis, the expected frequencies of the accidents on each of the days would be:

Days:	Sun.	Mon.	Tues.	Wed.	Thus.	Fri.	Sat.	
Total No. of Accidents	14	16	8	12	11	9	14	
No. of Accidents:	12	12	12	12	12	12	12	84

Therefore, $\chi^2 = \frac{(14-1\ )^2}{12} + \frac{(16-1\ )^2}{12} + \frac{(8-12)^2}{12} + \frac{(12-1\ )^2}{12} + \frac{(11-1\ )^2}{12} + \frac{(9-12)^2}{12} + \frac{(14-1\ )^2}{12}$

This implies that $\chi^2 = \frac{1}{12}(4 + 16 + 16 + 0 + 1 + 9 + 4) = \frac{50}{12} = 4.17$

The number of degrees of freedom = Number of observations- Number of independent constraints $= 7 - 1 = 6$. The tabulated $\chi^2_{0.05}$ for $6\, d.f. = 12.59$. Since the calculated χ^2 is much less than the tabulated value, it is highly insignificant and we accept the null hypothesis. Hence, we conclude that the accidents are uniformly distributed over the week.

***Example* 12.7.6**: The theory predicts the proportion of beans in the four groups A, B, C *and* D should be 9: 3: 3: 1. In an experiment among 1600 beans, the numbers in the four groups were 882, 313, 287 *and* 118. Does the experimental result support the theory?

Solution. Null Hypothesis: We set up the null hypothesis that the theory fits well into the experiment, i.e., the experimental results support the theory.

Under the null hypothesis, the expected (Theoretical) frequencies can be computed as follows:

Total number of beans $= 882 + 313 + 287 + 118 = 1600$

These are to be divided in the ratio 9: 3: 3: 1

$$\therefore\ E(882) = \frac{9}{16} \times 1600 = 900, \quad E(313) = \frac{3}{16} \times 1600 = 300$$

$$E(287) = \frac{3}{16} \times 1600 = 300, \quad E(118) = \frac{1}{16} \times 1600 = 100$$

$$\therefore \qquad \chi^2 = \sum \left[\frac{(O-E)^2}{E}\right]$$

$$= \frac{(882-900)^2}{900} + \frac{(313-300)^2}{300} + \frac{(287-300)^2}{300} + \frac{(118-100)^2}{100}$$

$$= 0.3600 + 0.5633 + 0.5633 + 3.2400 = 4.7266$$

$d.f. = 4 - 1 = 3$, And tabulated $\chi^2_{0.05}\ for\ 3\ d.f. = 7.815$

Since the calculated value of χ^2 is less than the tabulated value, it is not significant. Hence the null hypothesis may be accepted at 5% level of significance and we may conclude that there is good correspondence between theory and experiment.

***Example* 12.7.7**: $(2 \times 2$ *contingency table*$)$. For a 2×2 table,

a	b
c	d

Prove that chi-square test of independence gives

$$\chi^2 = \frac{(N(ad-bc)^2)}{(a+c)(b+d)(a+b)(c+d)},\ N = a+b+c+d \ldots (13.18)$$

Solution. Under the hypothesis of independence of attributes,

$E(a) = \frac{(a+b)(a+c)}{N}$, $E(b) = \frac{(a+b)(b+d)}{N}$, $E(c) = \frac{(a+c)(c+d)}{N}$

$E(d) = \frac{(b+d)(c+d)}{N}$

a	b	a+b
c	c	c+d
a+c	b+d	N

$$\therefore\ \chi^2 = \frac{[a-E(a)]^2}{E(a)} + \frac{[b-E(b)]^2}{E(b)} + \frac{[c-E(c)]^2}{E(c)} + \frac{[d-E(d)]^2}{E(d)} \quad \ldots (*)$$

$$a - E(a) = a - \frac{(a+b)(a+c)}{N}$$

$$= \frac{\big(a(a+b+c+d) - (a^2+ac+ab+bc)\big)}{N} = \frac{(ad-bc)}{N}$$

Similarly, we will get

$$b - E(b) = -\frac{(ad-bc)}{N} = c - E(c);\ d - E(d) = \frac{(ad-bc)}{N}$$

Substituting these values, we get

$$\chi^2 = \frac{(ad-bc)^2}{N^2}\left[\frac{1}{E(a)} + \frac{1}{E(b)} + \frac{1}{E(c)} + \frac{1}{E(d)}\right]$$

$$= \frac{(ad-bc)^2}{N^2}\left[\left\{\frac{1}{(a+b)(a+c)} + \frac{1}{(a+b)(b+d)}\right\} + \left\{\frac{1}{(a+c)(c+d)} + \frac{1}{(b+d)(c+d)}\right\}\right]$$

$$= \frac{(ad-bc)^2}{N^2}\left[\frac{b+d+a+c}{(a+b)(a+c)(b+d)} + \frac{b+d+a+c}{(a+c)(c+d)(b+d)}\right]$$

$$= (ad-bc)^2\left[\frac{c+d+a+b}{(a+b)(a+c)(b+d)(c+d)}\right]$$

$$= \frac{N(ad-bc)^2}{(a+b)(a+c)(b+d)(c+d)}$$

12.8 SUMMARY: -

This unit has presented about of chi square tests. In this unit we discussed M.G.F of Chi-square (χ^2) Distribution. Theorems on Chi-square (χ^2) distribution is also explaining in this unit. In this unit we also demonstrate the use of the chi-square distribution to conduct tests of (i) Goodness of fit, and (ii) Independence of attributes.

12.9 GLOSSARY:-

(i) Chi-Square Distribution: A kind of probability distribution, differentiated by Chi-Square Test their degree of freedom, used to test a number of different hypotheses about variances, proportions and distributional goodness of fit.
(ii) Expected Frequencies. The hypothetical data in the cells are called as expected frequencies.
(iii) Goodness of Fit. The chi-square test procedure used for the validation of our assumption about the probability distribution is called goodness of fit.
(iv)Observed Frequencies: The actual cell frequencies are called observed frequencies.

12.10. REFERENCES:-

1. S. C. Gupta and V. K. Kapoor: *Fundamentals of mathematical statistics*, Sultan Chand & Sons, 2020.
2. Seymour Lipschutz and John J. Schiller *: Schaum's Outline: Introduction to Probability and Statistics*, McGraw Hill Professional, 2017.
3. J. S. Milton and J. C. Arnold: *Introduction to Probability and Statistics* (4th Edition), Tata McGraw-Hill, 2003.
4. https://www.wikipedia.org.

12.11.SUGGESTED READINGS:-

1. Rohatgi, V. K., & Saleh, A. M. E. (2015). *An introduction to probability and statistics*. John Wiley & Sons
2. A.M. Goon: *Fundamental of Statistics* (7th Edition), 1998
3. R.V. Hogg and A.T. Craig: *Introduction to Mathematical Statistics*, MacMacMillan, 2002
4. R.V. Hogg, Joseph W. Mc Kean and T. Allen: Craig: *Introduction to Mathematical Statistics* (7th edition), Pearson Education, 2013

5. Irwin Miller and Marylees Miller John E. Freund: *Mathematical Statistics with Applications* (8th Edition). Pearson. Dorling Kindersley Pvt. Ltd. India, 2014

12.12 TERMINAL QUESTIONS:-

TQ1.A survey of 320 families with 5 children each revealed the following distribution:

No. of boys:	5	4	3	2	1	0
No. of girls:	0	1	2	3	4	5
No. of families:	14	56	110	88	40	12

Is this result consistent with the hypothesis that male and female births are equally probable?

TQ 2.A random sample of students of Bombay University was selected and asked their opinions about 'autonomous colleges'. The results are given below. The same number of each sex was included within each class-group. Test the hypothesis at 5% level that opinions are independent of the class groupings:

Class	Numbers		Total
	Favoring autonomous colleges	Opposed autonomous colleges	
First Yr. B.A/B.Sc/B.Com	120	80	200
First Yr. B.A/B.Sc/B.Com	130	70	200
First Yr. B.A/B.Sc/B.Com	70	30	100
M.A./ M.Sc./M.Com	80	20	100
Total	400	200	600

12.13 ANSWER

Answer of Check your progress Questions:-

CYQ 1: Independent
CYQ2: 5
CYQ3: Chi-Square test .

CYQ 4:Frequencies
CYQ 5 The degrees of freedom.

Answer of Terminal Questions:-

TQ 1. The null hypothesis of equal probability for male and female births is accepted.

TQ 2. We conclude that the opinions about autonomous colleges are dependent on the class-groupings.

UNIT 13:- TEST IN SAMPLING

CONTENTS:

13.1 INTRODUCTION:-

In previous unit we have defined Chi-Square Distribution and it's properties now in this unit we are explaining about t-distributions, F-distributions and z-distributions.

In statistics, the ***t***-distribution was first derived as a posterior distribution in ***1876*** by *Helmert* and *Lüroth*. The *t*-distribution also appeared in a more general form as Pearson Type ***IV*** distribution in *Karl Pearson's* ***1895*** paper. In the English-language literature, the distribution takes its name from *William Sealy Gosset's* ***1908*** paper in *Biometrika* under the pseudonym "Student". It became well known through the work of Ronald Fisher, who called the distribution "Student's distribution" and represented the test value with the letter *t*.

The *F* distribution was tabulated and the letter introduced by G. W. Snedecor. Fisher's z-Distribution was first described by *Ronald Fisher* in a paper delivered at the International Mathematical Congress of ***1924*** in *Toronto*. Nowadays one usually uses the *F*-distribution instead.

13.2. OBJECTIVES:-

After studying this unit learner will be able to:

1. Solve the problem related to t-distributions, F-Distribution and z-Distribution.
2. Discuss the important properties of t-distributions, F-Distribution and z-Distribution.
3. Analyze the application of t-distributions, F-Distribution and z-Distribution.
4. Explain the basic concept of t-distributions, F-Distribution and z-Distribution.

13.3. t -DISTRIBUTION(STUDENT'S):-

Let $x_i, (i = 1, 2, \ldots, n)$ be a random sample of size n from a normal population with mean μ and variance σ^2. Then Student's t is defined by the statistic

$$t = \frac{\bar{x}-\mu}{S/\sqrt{n}} \text{......................................(13.3.1)}$$

where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, is the sample mean and is an unbiased estimate of the population variance σ^2, and it follows Student's t-distribution with $v = (n-1)\ d.f.$ df. with probability density function.

$$f(t) = \frac{1}{\sqrt{v}\ B\left(\frac{1}{2}, \frac{v}{2}\right)} \cdot \frac{1}{\left[1+\frac{t^2}{v}\right]^{(v+1)/2}}; -\infty < t < \infty \ldots\ldots\ldots (\mathbf{13.3.2})$$

Remark 13.3.1. A statistic t following Student's t -distribution with $d.f.$ will be abbreviated as $t \sim t_n$.

Remark 13.3.2. If we take $v = 0$ in (13.5.2), we get

$$f(t) = \frac{1}{B\left(\frac{1}{2}, \frac{1}{2}\right)} \cdot \frac{1}{(1+t^2)} = \frac{1}{\pi} \cdot \frac{1}{(1+t^2)}; -\infty < t < \infty \qquad \left[\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}\right]$$

which is the $p.d.f.$ of standard Cauchy distribution. Hence $v = 1$, Student's t-distribution reduces to Cauchy distribution.

Applications of t -distribution The t -distribution has a wide number of applications in Statistics, some of which are enumerated below.

(i) To test if the sample mean $(\bar{x})$ differs significantly from the hypothetical value μ of the population mean.

(ii) To test the significance of the difference between two sample means.

(iii) To test the significance of an observed sample correlation co-efficient and sample regression coefficient.

(iv) To test the significance of observed partial and multiple correlation coefficients. In the following sections we will discuss these applications in detail, one by one.

t-test for Single Mean. Suppose we want to test :

i. if a random sample $x_i, (i = 1, 2, \ldots, n)$ of size n has been drawn from normal population with a specified mean, say μ_0, or

ii. if the sample mean differs significantly from the hypothetical value$H_\circ$of the population mean.

Under the null hypothesis H_0:

(i) The sample has been drawn from the population with mean μ_0 or

(ii) There is no significant difference between the sample mean $\bar{x}$ and the population mean μ_0,

the statistic $t = \frac{\bar{x}-\mu}{\sqrt{s^2/n}}$ **……………………………….(13.5.3)**

where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and

$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2, \ldots\ldots\ldots\ldots\ldots\ldots. \mathbf{(13.5.4)}$

follows Sudent's t –distribution with $(n-1)d.f.$

We now compare the calculated value of t with the tabulated value at certain level of significance. If calculated $|t| >$ tabulated t, null hypothesis is rejected and if calculated $|t| >$ tabulated t, H_0 may be accepted at the level of significance adopted.

Assumptions for Student's t-test. The following assumptions are made in the Student's t-test:

(i) The parent population from which the sample is drawn is normal.

(ii) The sample observations are independent, *i.e.,* the sample is random.

(iii) The population standard deviation σ is unknown.

Paired t-test For Difference of Means. Let us now consider the case when (i) the sample sizes are equal, *i.e.* ,$n_1 = n_2 = n$ (say), and (ii) the two samples are not independent but the sample observations are paired together, *i.e.*, the pair of observations (x_i, y_i). $(i = 1, 2, \ldots, n)$ corresponds to the same $(i$th) sample unit. The problem is to test if the sample means differ significantly or not.

For example, suppose we want to test the efficacy of a particular drug, say, for inducing sleep. Let $x_i$andy_i, $(i = 1, 2, \ldots, n)$ be the readings, in hours of sleep, on the ith individual, before and after the drug is given respectively. Here instead of applying the difference of the means test discussed in above, we apply the paired t-test given below.

Here we consider the increments, $d_i = x_i - y_i$, $(i = 1, 2, \ldots, n)$.

Under the null hypothesis, H_0 that increments are due to fluctuations of sampling, *i.e.,* the drug is not responsible for these increments, the statistic.

$$t = \frac{\bar{d}}{s/\sqrt{n}}$$

Where $\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i$ and $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(d_i - \bar{d}\right)^2$. Follows Student's t –distribution with $(n-1)d.f.$

Check Your Progress

1. The population ……………………….is unknown in $t -$ test.
2. The sample observations are ……………….

13.4.F-DISTRIBUTION:-

If X and Y are two independent chi-square variates with v_1 *and* v_2 d.f. respectively, then $F -$ statistic is defined by

$$F = \frac{\frac{X}{v_1}}{\frac{Y}{v_2}}$$

In other words, F is defined as the ratio of two independent chi-square variates divided by their respective degrees of freedom and it follows Snedecor's F-distribution with (v_1, v_2) d.f. with probability function given by

Remarks 13.4.1. The sampling distribution of $F -$ statistic does not involve any population parameters and depends only on the degrees of freedomv_1 *and* v_2.

A statistic F following Snedecor's F-distribution with $(v_1, v_2)\, d.f.$ will be denoted by $F \sim F(v_1, v_2)$.

Application of *F*- distribution:

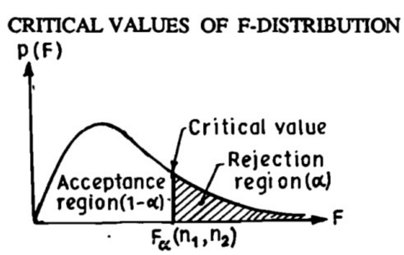
***F*-test for Equality of Population Variances.** Suppose we want to test (i) whether two independent samples $x_i,\ (i = 1,2,\ldots,n_1)\ and\ y_{j,}(j = 1,2,\ldots,n_2)$ have been drawn from the normal populations with the same variance σ^2 , (say), or (ii) Two independent estimates of the population variance are homogeneous, the statistic F is given by

$$F = \frac{S_x^2}{S_y^2}$$

Where $S_X^2 = \frac{1}{n_1-1}\sum_{i=1}^{n_1}(x_i - \overline{x})^2\ and\ S_Y^2 = \frac{1}{n_2-1}\sum_{i=1}^{n_2}(y_i - \overline{y})^2$
Are unbiased estimates of the common population variance σ^2 obtained from two independent samples an dit follows Snedecor's F-distribution with $(v_1, v_2)\ d.f.$[where[$v_1 = n_1 - 1\ and\ v_2 = n_2 - 1$].

Remark 13.4.2. In $F = \frac{S_x^2}{S_y^2}$greater of the two variances $S_{x^2}\ and\ S_{y^2}$ is to be taken in the numerator and n_1 corresponds to the greater variance. By comparing the calculated value of F obtained by using $F = \frac{S_x^2}{S_y^2}$for the two given samples with the tabulated value of F for $(n_{1,}n_2)$ d.f. at certain level of significance$(5\%\ or\ 1\%)$, H_0 is either rejected or accepted.

***Critical values of F*- distribution.** The available F-table (given in the Appendix at the end of the book) give the critical values of F for the right tailed test, I.e., the critical region is determined by the right-tail areas. Thus, the significant value $F_\alpha(n_1, n_2)$ at level of significance α and $(n_1, n_2)d.f.$ is determined by
$P[F > F_\alpha(n_1, n_2)] = \alpha,$



Remark 13.4.3 Z-distribution tends to normal distribution with mean $\frac{1}{2}\left(\frac{1}{v_1} - \frac{1}{v_2}\right)$ and variance $\frac{1}{2}\left(\frac{1}{v_1} + \frac{1}{v_2}\right)$, as v_1 and v_2 become large.

Relation Between t-Distribution and *F*- distribution. A relation is derived between the percentile points of a t-distribution with n degrees of freedom and those of an F-distribution with n and n degrees of freedom. In effect, the t-percentiles can be obtained by a simple transformation from the "diagonal" entries of an F-table.

Relation Between F-Distribution and Chi-square (χ^2) Distribution

In $F(v_1, v_2)$.distribution if we get $v_2 \longrightarrow \infty$, then $\chi^2 = v_1 F(v_1, v_2)$ folllows Chi-square (χ^2) distribution with v_1 degree of freedom.Both the F-distribution and the chi-square distribution are positively skewed distributions.

Check your Progress

3.F-Distribution is defined as the ratio of two independent …………….

4. Which of the following distributions is Continuous?

a) Binomial Distribution

b) Hyper-geometric Distribution

c) F-Distribution

d) Poisson Distribution

5. F-Distribution cannot take negative values. True\False

13.5 FISHER'S Z –TRANSFORMATION:-

To test the significance of an observed sample correlation coefficient from an uncorrelated bivariate normal population, t-testis used. But in random sample of size n, from a bivariate normal population in which $\rho \neq 0$, Prof. R.A. Fisher proved that the distribution of $'r'$ is by no means normal and in the neighborhood of $\rho = \pm 1$, its probability curve is extremely skewed even for large n. If $\rho \neq 0$, Fisher suggested the following transformation

$$Z = \frac{1}{2} Log_e \left(\frac{1+r}{1-r}\right) = tanh^{-1} r$$

and proved that even for small samples, the distribution of Z is approximately normal with mean

$$\xi = \frac{1}{2} \log_e \frac{1+\rho}{1-\rho} = \tanh^{-1}(\rho) \qquad \ldots$$

and variance $\frac{1}{(n-3)}$ and for large values of n, say>50, the approximation is fairly good.

Application of z-transformation:

To test if an observed value of 'r' differs significantly from a hypothetical value ρ of the population correlation coefficient.
H_0: There is no significant difference between $r\ and\ \rho$. In other words, the given sample has been drawn from a bivariate normal population with correlation coefficient ρ.
If we take $Z = \frac{1}{2} Log_e \left(\frac{1+r}{1-r}\right)$ and $\xi = \frac{1}{2} \log_e \frac{1+\rho}{1-\rho}$.
Then under $H_0, Z \sim N\left(\xi, \frac{1}{n-3}\right)$
i.e. $\frac{(z-\xi)}{\sqrt{\frac{1}{n-3}}} \sim N(0,1)$.
Thus if $(Z - \xi)\sqrt{n-3} > 1.96$, H_0 is rejected at 5% level of significance and if it is greater than 2.58, H_0 is rejected at 1% level of significance.

13.6. SOLVED EXAMPLES:-

Example 13.6.1. A machinist is making engine parts with axle diameters of 0.700 inch. A random sample, of 10 parts shows a mean diameter of 0.742 inch with a standard deviation of 0.040 inch. Compute the statistic you would use to test whether the work is meeting the specifications. Also state how you would proceed further.

Solution. Here we are given :
μ=0. 700 inches, $\bar{x}$= 0.742 inches, s = 0-040 inches and n = 10

Null Hypothesis:
$H_0: \mu = 0.700,\ i.e.\ the\ product\ is\ confirming\ to\ specifications.$
Alternative Hypothesis, $H_1: \mu \neq 0.700,$
Test Statistic. Under H_0, the test statistic is:

$$t = \frac{\bar{x} - \mu}{\sqrt{s^2/n}} = \frac{\bar{x} - \mu}{\sqrt{s^2/(n-1)}} \sim t_{(n-1)}$$

$$t = \frac{\sqrt{9}(0.742 - 0.700)}{0.040} = 3.15$$

Now degree of freedom $(d.f.) = 10 - 1 = 9$. We will now compare this calculated value with tabulated value at $9(d.f.)$ and certain level of significance say 5%. Let this tabulated value be denoted by t_0.

(i) If calculated $'t', viz., 3.15 > t_0$. It implies that the value of t is significant. Therefore $\bar{x}$ differs significantly from $\bar{x}$ and $\mu. H_0$ is rejected at this level of significance and we conclude that the product is not meeting the specifications.

(ii) If calculated $'t', viz., 3.15 < t_0$. It implies that the value of t is not significant. Therefore $\bar{x}$ no significant difference between $\bar{x}$ and μ. and H_0 is accepted at this level of significance and we may take the product conforming to specifications.

Example 13.6.2. The mean weekly sales of soap bars in departmental stores was 146.3 bars per store. After an advertising campaign the mean weekly sales in 22 stores for a typical week increased to 153.7 and showed a standard deviation of 17.2. Was the advertising campaign successful?

Solution. We are given: $n = 22, \quad \bar{x} = 153.7, \; s = 17.2,$
Null Hypothesis. The advertising campaign is not successful, i.e..
H_0: $\mu = 146.3$
Alternative Hypothesis. H_1: μ> 146-3.(Right-tail).
Test Statistic. Under the null hypothesis, the test statistic is:

$$t = \frac{\bar{x} - \mu}{\sqrt{s^2/(n-1)}} \sim t_{22-1} = t_{21}$$

Now $$t = \frac{153.7 - 146.3}{\sqrt{(17.2)^2/21}} = \frac{7.4 \times \sqrt{21}}{17.2} = 9.03$$

Conclusion. Tabulated value of t for 21 df. at 5% level of significance for single-tailed test is 1.72. Since calculated value is much greater than the tabulated value, it is highly significant. Hence we reject the null hypothesis and conclude that the advertising campaign was definitely successful in promoting sales.

Example 13.6.4. Samples of two types of electric light bulbs were tested for length of life and following data were obtained:

	Type I	***Type II***
Sample No	$n_1 = 8$	$n_2 = 7$
Sample Means	$\bar{x}_1 = 1234\ hrs.$	$\bar{x}_2 = 1{,}036\ hrs.$
Sample S.D.'s	$s_1 = 36\ hrs.$	$s_2 = 40\ hrs.$

Is the difference in the means sufficient to warrant that type I is superior to type II regarding length of life?

Solution. Null Hypothesis, $H_0: \mu_x = \mu_y, i.e$ ..., the two types I and II of electric bulbs are identical.

Alternative Hypothesis.$H_1: \mu_x > \mu_y, i.e$ ...type I is superior to type II,
Test Statistic. Under H_0, The test statistic is:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{S^2\left(\frac{1}{n_1} - \frac{1}{n_2}\right)}} \sim t_{n_1+n_2-2} = t_{13,}$$

Where, $S^2 = \frac{1}{n_1+n_2-2} = [\sum(x_1 - \overline{x_1})^2 + \sum(x_2 - \overline{x_2})^2$

$$= \frac{1}{n_1 + n_2 - 2}[n_1 {s_1}^2 + n_2 {s_2}^2] = \frac{1}{13}[8 \times (36)^2 + 7 \times 40^2]$$
$$= 1659.08$$
$$\therefore \qquad t = \frac{1234 - 1036}{\sqrt{1659.08\left(\frac{1}{8} + \frac{1}{7}\right)}} = \frac{198}{\sqrt{1659.08 \times 0.2679}} = 9.39$$

Tabulated value of t for 13 $d.f.$ at 5% level of significance for right (single) tailed test is 1·77.[This is the value of $t_{0.10}$ for 13 $d.f.$ From two tail tables given in Appendix].

Conclusion. Since Calculated *'t'* is much greater than tabulated *'t'*, it is highly significant and H_0 is rejected. Hence the two types of electric bulbs differ significantly. Further since $\overline{x_1}$ is much greater than $\overline{x_2}$, we conclude that type I is definitely superior to type II.

Example 13.6.5. A certain stimulus administered to each of the 12 patients resulted in the
following increase of blood pressure :
5, 2, 8, -1, 3,0,-2, 1,5, 0, 4 and 6. Can it be calculated that the stimulus will, in general, be accompanied by an increase in blood pressure?

Solution. Here we are given the increments in blood pressure *i.e.,*
$d_i (= x_i, y_i)$.
Null Hypothesis,H_0: $\mu_x = \mu_y$, $i.e.$, there is no significant difference in the blood pressure readings of the patients before and after the drug. In other words, the given increments are just by chance (fluctuations of sampling) and not due to the stimulus.

Alternative Hypothesis, $H_1 : \mu_x < \mu_y$, $i.e.$, the stimulus results in an increase in blood pressure.

Test statistic. Under H_0, the test statisti is :

$$t = \frac{\bar{d}}{s/\sqrt{n}} \sim t_{(n-1)}$$

$$S^2 = \frac{1}{n-1}\sum (d - \bar{d})^2 = \frac{1}{n-1}[\sum d^2 - \frac{(\sum d)^2}{n}]$$
$$= \frac{1}{11}\left[185 - \frac{(31)^2}{12}\right] = \frac{1}{11}(185 - 80.08) = 9.5382$$

And $\qquad \bar{d} = \frac{\sum d}{n} = \frac{31}{12} = 2.58$

$$\therefore \qquad t = \frac{\bar{d}}{s/\sqrt{n}} = \frac{2.58 \times \sqrt{12}}{\sqrt{9.5382}} = \frac{2.58 \times 3.464}{3.09} = 2.89$$

Tabulated $t_{0.05}$for $11 d.f.$ for right-tail test is 1.80.
[This is the value of $t_{0.05}$for $11 d.f.$ in the Table for two-tailed test given in the Appendix]

Example 13.6.6. Pumpkins were grown under two experimental conditions. Two random samples of $11\ and\ 9$ pumpkins show the sample standard deviations of their weights as $0.8\ and\ 0.5$ respectively. Assuming that the weight distributions are normal, test the hypothesis that the true variances are equal, against the alternative that they are not, at the 10% level.[Assume that $\left[P\left(F_{10,8} \geq 3.35\right) = 0.05\ and\ P\left(F_{8,10} \geq 3.07\right) = 0.05.\right]$

Solution. We want to test Null Hypothesis, $H_0: \sigma_x^2 = \sigma_r^2$. against the Alternative Hypothesis,
$H_1: \sigma_x^2 \neq \sigma_r^2$.(Two-tailed).

We are given:

$$n_1 = 11,\ n_2 = 9,\ s_x = 0.8\ and\ s_y = 0.5.$$

Under the null Hypothesis, $H_0: \sigma_x = \sigma_r$. the statistic

$$F = \frac{S_{X^2}}{S_{Y^2}}$$

Follows F-distribution with $(n_1 - 1, n_2 - 1) d.f.$

Now $$n_1 s_{X^2} = (n_1 - 1) S_{X^2}$$

$$\therefore\ S_{X^2} = \left(\frac{n_1}{n_1 - 1}\right) s_{X^2} = \left(\frac{11}{10}\right) \times (0.8)^2 = 0.704$$

Similarly, $$S_{Y^2} = \left(\frac{n_1}{n_1 - 1}\right) s_{Y^2} = \left(\frac{9}{8}\right) \times (0.5)^2 = 0.28125$$

$$\therefore\ F = \frac{0.704}{0.28125} = 2.5$$

The significant values of F for two tailed test at level of significance $\alpha = 0.10$ are:

$$F > F_{10,8}\left(\frac{\alpha}{2}\right) = F_{10,8}(0.05)\ and\ F > F_{10,8}\left(1 - \frac{\alpha}{2}\right) = F_{10,8}(0.95) \quad \ldots\ldots\ldots\mathbf{(13.6.6.1)}$$

We are given the tabulated (significant) values:

$$P\left[F_{10,8} \geq 3.35\right] = 0.05\ \Rightarrow\ F_{10,8}(0.05) = 3.35 \quad \ldots\ldots\ldots\ldots\mathbf{(13.6.6.2)}$$

Also $$P\left[F_{8,10} \geq 3.07\right] = 0.05\ \Rightarrow\ P\left[\frac{1}{F_{8,10}} \leq \frac{1}{3.07}\right] = 0.05$$

$$\Rightarrow P\left[F_{10,8} \leq 0.326\right] = 0.05\ \Rightarrow\ P\left[F_{10,8} \geq 0.326\right] = 0.95 \quad \ldots\ldots\ldots\ldots\mathbf{(13.6.6.3)}$$

Hence from $(\mathbf{13.6.6.1})$, $(\mathbf{13.6.6.2})$ and $(\mathbf{13.6.6.3})$, the critical values for testing H_0: $\sigma_x^2 = \sigma_y^2,\ against\ H_1: \sigma_x^2 \neq \sigma_y^2$. At level of significance $\alpha = 0.10$ are given by: $F > 3.35\ and\ F < 0.326 = 0.33$ .Since, the calculated value of $F(= 2.5)$ lies between $0.33\ and\ 3.35$, it is not significant and hence null hypothesis of equality of population variances may be accepted at level of significance $\alpha = 0.10$.

Remark 13.6.7Z-distribution tends to normal distribution with mean $\frac{1}{2}\left(\frac{1}{v_1}-\frac{1}{v_2}\right)$ and variance $\frac{1}{2}\left(\frac{1}{v_1}+\frac{1}{v_2}\right)$, as v_1 and v_2 become large.

Example 13.6.8.A correlation coefficient of 0.72 is obtained from a sample of 29 pairs of observations.

(i) Can the sample be regarded as drawn from a bivariate normal population in which true correlation coefficient is 0.8?
(ii) Obtain 95% confidence limits for ρ in the light of the information provided by the sample.

Solution. (i) H_0**:** There is no, significant difference between $r = 0.72;\ and\ \rho = 0.80, i.e.,$ the sample can be regarded as drawn from the bivariate normal population with $\rho = 0.8$.

Here $Z = \frac{1}{2} Log_e\left(\frac{1+r}{1-r}\right) = 1.1513 \log_{10}\left(\frac{1+r}{1-r}\right)$

$= 1.1513\ \log_{10} 6.14 = 0.907$

$$\xi = \frac{1}{2}\log_e \frac{1+\rho}{1-\rho} = 1.1513\ \log_{10}\left(\frac{1+0.8}{1-0.8}\right) = 1.1513 \times 0.9541$$

$= 1.1$

S.E.(Z)=$\frac{1}{\sqrt{n-3}} = \frac{1}{\sqrt{26}} = 0.196$

Under H_0, the test statistic is $U = \frac{Z-\xi}{\frac{1}{\sqrt{n-3}}} \sim N(0,1)$

$$U = \frac{0.907 - 1.100}{0.196} = -0.985$$

Since $|U| < 1.96$, it is not significant at 5% level of significance and H_0 may be accepted. Hence the sample may be regarded as coming from a bivariate normal population with $\rho = 0.8$.

(iii)95% Confidence limits for ρ on the basis of the information supplied by the sample, are given by

$|U| \leq 1.96$

$$|Z - \xi| \leq 1.96 \times \frac{1}{\sqrt{n-3}} = 1.96 \times 0.196$$

$$\Rightarrow |0.907 - \xi| \leq 0.384 \Rightarrow 0.907 - 0.384 \leq \xi \leq 0.907 + 0.384$$

$$\Rightarrow 0.523 \leq \xi \leq 1.291$$

$$\Rightarrow 0.523 \leq \frac{1}{2}\log_e\left(\frac{1+\rho}{1+\rho}\right) \leq 1.291$$

$$\Rightarrow 0.523 \leq 1.1513 \log_{10}\left(\frac{1+\rho}{1+\rho}\right) \leq 1.291$$

$$\Rightarrow \frac{0.523}{1.1513} \leq \log_{10}\left(\frac{1+\rho}{1+\rho}\right) \leq \frac{1.291}{1.1513}$$

$\Rightarrow 0.4543 \leq \log_{10}\left(\frac{1+\rho}{1+\rho}\right) \leq 1.1213.$

Now $\log_{10}\left(\frac{1+\rho}{1+\rho}\right) = 0.4543$ *and* $\log_{10}\left(\frac{1+\rho}{1+\rho}\right) = 1.1213 \Rightarrow \left(\frac{1+\rho}{1+\rho}\right) = Anti\log(0.4543) = 2.846 \Rightarrow \frac{1+\rho}{1+\rho} = Anti\log(1.1213) = 13.22. \Rightarrow \rho = \frac{2.846-1}{2.846+1} = \frac{1.846}{3.846} = 0.4799. \Rightarrow \rho = \frac{13.22-1}{13.22+1} = \frac{12.22}{14.22} = 0.86$

13.7.SUMMARY: -

In this unit we have explained the following three specific distributions i) t-distribution ii) F-distribution iii) z-distribution and explained the use of these distributions as sampling distribution.

13.8. GLOSSARY:-

1. t-distributions
2. F-distribution
3. z-dis tribution
4. random sample
5. mean
6. Variance
7. Statistic

13.9.REFERENCES:-

1. S. C. Gupta and V. K. Kapoor: *Fundamentals of mathematical statistics*, Sultan Chand & Sons, 2020.
2. Seymour Lipschutz and John J. Schiller *:Schaum's Outline: Introduction to Probability and Statistics*, McGraw Hill Professional, 2017.
3. J. S. Milton and J. C. Arnold: *Introduction to Probability and Statistics* (4th Edition), Tata McGraw-Hill, 2003.
4. https://www.wikipedia.org.

13.10.SUGGESTED READINGS:-

1. Rohatgi, V. K., &Saleh, A. M. E. (2015). *An introduction to probability and statistics*. John Wiley & Sons
2. A.M. Goon: *Fundamental of Statistics* (7th Edition), 1998
3. R.V. Hogg and A.T. Craig: *Introduction to Mathematical Statistics*, MacMacMillan, 2002

4. R.V. Hogg, Joseph W. Mc Kean and T. Allen: Craig: *Introduction to Mathematical Statistics* (7th edition), Pearson Education, 2013
5. Irwin Miller and Marylees Miller John E. Freund: *Mathematical Statistics with Applications* (8th Edition). Pearson. Dorling Kindersley Pvt. Ltd. India, 2014

13.11.TERMINAL QUESTIONS:-

TQ 13.11.1 The heights of 10 males of a given locality are found tobe 70, 67, 62, 68, 61, 68, 70, 64, 64, 66 inches. Is it reasonable to believe thatthe average height is greater than 64 inches? Test at 5% significance level, assuming thatfor9 degrees offreedom P (t > 1.83) = 0·05.

TQ 13.11.2 The heights of six randomly chosen sailors are ininches: 63, 65, 68, 69, 71 and 72. Those of 10 randomly chosen soldiers are61, 62, 65, 66, 69, 69, 70, 71, 72 and 73. Discuss the light that these datathrow on the suggestion that sailors are on theaverage taller than soldiers.

TQ13.11.3 Below are given the gain in weight (in lbs.) of pigs fed on two diets A and B.
Gain in weight
Diet A : 25, 32, 30, 34, 24, 14, 32, 24, 30, 35, 25
Diet B : 44, 34, 22, 10, 47, 31, 40, 30, 32, 35, 18, 21, 35, 29, 22
Test, if the two diets differ significantly as regards their effect on increase in weight

TQ 13.11.4 In one sample of 8 observations, the sum of the squaresof deviations of the sample values from the sample mean was84.4 and in theother sample of 10 observations it was 102.6. Test whether this difference issignificant at 5 percent level, given that the 5 percent point of F for $n_1 = 7$ and$n_2 = 9$ degrees offreedom is 3·29.

TQ 13.11.5 Two random samples gave the following results:

Sample	Size	Sample mean	Sum of squares of deviations from the mean
1	10	15	90
2	12	14	108

Test whether the samples come from the same normal population at 5% level of significance.
{Given:
$F_{0.05}(9,11) = 2.90, F_{0.05}(11,9) = 3.10\ (approx.), t_{0.05}(20) = 2.086\ and\ t_{0.05}(22) = 2.07$ }.

13.12 ANSWER:-

Answer of Check your progress:-

CYQ1. Standard deviation σ.
CYQ2. Independent.
CYQ3. chi-squarevariates.
CHQ4. F-Distribution
CHQ5. True.

Answer of Terminal Questions:-

TQ 13.11.1. We conclude that average height is greater than 64 inches.
TQ 13.11.2. At 5% level of significance and we conclude that the data are inconsistent with the suggestion that the sailors are on the average taller than soldiers.
TQ 13.11.3. Since calculated |t| is less than tabulated t, H_0 may be accepted at 5% level of significance and we may conclude that the two diets do not difer significantly, as regards their effect on increase in weight.
TQ 13.11.4. Since calculated $F < F_{0.05}$, H_0may be accepted at 5% level of significance.
TQ 13.11.5. The given samples have been drawn from the same normal population.

UNIT 14:- THEORY OF ESTIMATOR

CONTENTS:

14.1.INTRODUCTION:-

The theory of estimation was founded by Prof. R.A. In this unit of our course, we will learn about the theory of estimator. Estimation theory is a branch of statistics that deals with estimating the values of parameters based on measured empirical data that has a random component. The parameters describe an underlying physical setting in such a way that their value affects the distribution of the measured data. In this unit we are explaining about estimation theory.

Prof. R.A. Fisher
(1890-1962)
Fig:14.1.1
Ref:
https://en.wikipedia.org/wiki/Ronald_Fisher#/media/File:Youngronaldfisher2.JPG

14.2.OBJECTIVES:-

After studying this unit learner will be able to:

1. Know more about point estimator.
2. Understand the properties of estimator.
3. Choose relatively best estimator among some estimator of same parameter.

14.3.POINT ESTIMATOR:-

Let X be random variable describe the population under the study. i.e., either X has a p.d.f. or p.m.f. The distribution of X may depend on some unknown parameter ϴ. Generally we have the case that distribution of X is known but its parameter is unknown.

For example we have a population random variable X, and we do have information that this X follows normal distribution (μ, ***), but parameter μ and *** are unknown. Now if we estimate these two parameter μ and *** by picking some sample. Then we can approximate the information about the given population.

Consider a sample of size n. Let $x_1, x_2, \dots, x_n$ be sample values. And each $x_i's$ are independent and identically distributed random variables, which has the same distribution as that of X. i.e. if p.d.f. of X is g(x). Then p.d.f. of statistic $t = t(x_1, x_2, \dots, x_n)$ is $f(x_1, x_2, \dots, x_n) = g(x_1)g(x_2)\dots g(x_n) = \big(g(x)\big)^n$.

The above distribution is called sampling distribution of t.We already know that any function $x_1, x_2, \dots, x_n$ is a statistic $t = t(x_1, x_2, \dots, x_n)$. In previous unit we saw two statistics $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$and $S^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$. Some other examples are $t_1 = max(x_1, x_2, \dots, x_n)$ and $t_2 = min(x_1, x_2, \dots, x_n)$.

If we use any statistic t to estimate the unknown parameter θ, it is called point estimator of θ.

Now suppose from a sample we have two statistics t_1 andt_2 for the population parameter θ. Then question is: Among the statistics $t_1$andt_2, which one is better estimator for θ? Obviously, as a solution the first criteria comes in mind is that $t_1$is preferrable if modulus of difference $|t_1 - \theta|$ is less than$|t_2 - \theta|$. But in most of the case is unknown. Thus we require some other criterion for selecting better estimator.

14.4. CHARACTERISTICS OF ESTIMATOR:-

The following are some of the criteria that should be satisfied by a good estimator:

(i) Unbiasedness
(ii) Consistency
(iii) Efficiency
(iv) Sufficiency

14.4.1.UNBIASED ESTIMATOR:-

An estimator $T_n = T(x_1, x_{2,}, \dots \dots., x_n)$ is said to be an unbiased estimator of $\gamma(\theta)$ if

$$E(T_n) = \gamma(\theta), \text{ for all } \theta \in \Theta.$$

We have already seen that sample mean is an unbiased estimator of population mean.

Example Show that S^2 which is defined as follows:

$$S^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}.$$

is an unbiased estimator of population variance σ^2.

Solution: We know for a random variable Y:

$$V(Y) = E(Y^2) - E(Y)^2$$

Therefore

$$E(Y^2) = V(Y) + E(Y)^2.$$

Consider a sample with n units $x_1, x_2, \dots, x_n$. Then we have already seen that for each i=1,2,…,n$E(x_i) = \mu$, $V(x_i) = \sigma^2, E(\bar{x}) = \mu$, and$V(\bar{x}) = \frac{\sigma^2}{n}$. Thus

$$E({x_i}^2) = V(x_i) + E(x_i)^2 = \sigma^2 + \mu^2.$$

And

$$E(\bar{x}^2) = V(\bar{x}) + E(\bar{x})^2 = \frac{\sigma^2}{n} + \mu^2$$

Now, given statistic is S^2:

$$S^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}.$$

Then

$$E(S^2) = E\left(\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}\right) = \frac{1}{n-1} E\left(\sum_{i=1}^{n}(x_i - \bar{x})^2\right)$$

This implies that

$$E(S^2) = \frac{1}{n-1} \sum_{i=1}^{n}[E(x_i^2) - E(\bar{x}^2)]$$

$$= \frac{1}{n-1} \sum_{i=1}^{n}[\sigma^2 + \mu^2 - \frac{\sigma^2}{n} + \mu^2].$$

Consequently,

$$E(S^2) = \frac{1}{n-1}\sum_{i=1}^{n}\left[\sigma^2 - \frac{\sigma^2}{n}\right] = \frac{1}{n-1}\sum_{i=1}^{n}\left[\frac{(n-1)\sigma^2}{n}\right] = \sigma^2.$$

Thus$S^2$is an unbiased estimator of population variance σ^2.
The following table is list of expectations of some important statistics:

S.N.	t	E(t)
1	x_i	$E(x_i) = \mu$
2	$\bar{x}$	$E(\bar{x}) = \mu$
3	${x_i}^2$	$E({x_i}^2) = \sigma^2 + \mu^2$.
4	$\bar{x}^2$	$E(\bar{x}^2) = \frac{\sigma^2}{n} + \mu^2$
5	$x_i\bar{x}$	$E(x_i\bar{x}) = \frac{\sigma^2}{n} + \mu^2$
6	S^2	$E(S^2) = \sigma^2$

Remark:

1. Unbiased estimator may not be exist for a parameter.
2. There are more than one parameter for given parameter.
3. If t_1 and t_2 are unbiased estimator of Θ, then for any $0 \le k \le 1, t = kt_1 + (1-k)t_2$ is also an unbiased estimator of Θ.

14.4.2.CONSISTENCY:-

An estimator $t_n = t(x_1, x_2, \dots, x_n)$ based on a sample of size n is said to be consistent estimator of Θ, if the sequence of statistic, t_n converges to Θ in probability. i.e. we say that t_n is a consistent estimator of Θ, if for every $\varepsilon > 0$ and$\delta > 0$ there exists a natural number $n_0(\varepsilon, \delta)$ such that
$P[|t_n - \Theta| < \varepsilon] > 1 - \delta; \quad \forall n \ge n_0$.In other words,
$P[|t_n - \Theta| < \varepsilon] \to 1$ as $n \to \infty$.
As we can see in above definition, consistent estimator will come more closer to Θ, if the sample size increases.

Example: For each natural number n, consider $x_1, x_2, \dots, x_n$ be sample values. For each natural number n, let us denote sample mean by $(\bar{x})_n = \frac{\sum_{i=1}^{n} x_i}{n}$. Then by weak law of large numbers, $(\bar{x})_n$ converge to μ (population mean). Hence sample mean is a consistent estimator of the population.

Theorem 14.4.2.1 (Invariance property of consistent estimators): If t_n is a consistent estimator of θ and $g(\theta)$ be continuous function of θ, then $g(t_n)$ is a consistent estimator of $g(\theta)$.

14.4.3.EFFICIENT ESTIMATOR AND EFFICIENCY:-

Suppose to estimate a parameter, we have two statistics, both are unbiased and consistent. Yes, This case may happen. For example, in sampling from a normal population N(μ, σ^2), when σ^2 is known, sample mean $\bar{x}$ is an unbiased and consistent estimator of μ. Therefore, we need some more criteria to choose a better estimator between the estimators which are consistent. This criterion is known as efficiency. The efficiency of estimators is a comparative analysis which is based on variance of estimators of the sampling distribution. Suppose that, for a parameter, we have two consistent estimatorst_1 and $t_2$with following properties:

$$Var(t_1) < Var(t_2), \text{for all sample size.}$$

Then we say that t_1 is more efficient that t_2.

For a sampling from a normal population, it is known that $Var(\bar{x}) = \frac{\sigma^2}{n}$and $Var(Med) = 1.57\frac{\sigma^2}{n}$. Hence for normal distribution, sample mean is more efficient estimator than the sample median for μ.

Most Efficient Estimator. Suppose we have set of consistent estimators for a parameter, and there exists an estimator whose sampling variance is less than to other estimator from that set. Then this estimator is called the most efficient estimator among them.

Efficiency If$t^*$is the most efficient estimator with variance V^* and t is any other estimator with variance V (with $V \neq 0$), then the efficiency $E(t)$ of $t$is defined as:

$$E(t) = \frac{V^*}{V}.$$

Remark: For a sampling from a normal population $\bar{x}$ is the most efficient estimator of μ.

As we seen earlier, for a sampling from a normal population, $Var(\bar{x}) = \frac{\sigma^2}{n}$and $Var(Med) = \pi\,\sigma^2/(2n)$. Since $\bar{x}$ is the most efficient estimator of μ. Therefore

$$E(Med) = \frac{Var(\bar{x})}{Var(Med)} = \frac{\sigma^2/n}{\pi\,\sigma^2/(2n)} = \frac{2}{\pi} = 0\cdot 637$$

Example: Let t be an estimator. Then show that efficiency $0 \leq E(t) \leq 1$.

Solution: Let V be the variance of estimator t (with $V \neq 0$) and t^* is the most efficient estimator with variance V^* . Since t is the most efficient estimator, therefore $V^* \leq V$. Also, both V^* and V are positive therefore $0 \leq E(t) \leq 1$.

Minimum Variance Unbiased (M.V.U.) Estimator.

Consider a statistic $t = t(x_1, x_2, \ldots\ldots x_n)$, based on sample of size n with following two properties:

i. $t$is unbiased for θ.
ii. t has the smallest variance among the set of all unbiased estimators of θ.

Then t is called the minimum variance unbiased estimator (MVUE) of θ.

The following theorem tell us about uniqueness properties of MVU estimators.

Theorem14.4.3.1 . An M.V.U. is unique in the sense that if $t_1$and t_2 are MVU estimators for θ, then $t_1 = t_2$, almost surely.

Proof. We are given that

$$E(t_1) = E(t_2) = \theta$$

And $\quad Var(t_1) = Var(t_2)$

Consider a new estimator, $t = \frac{1}{2}(t_1 + t_2)$

Then

$$E(t) = \frac{1}{2}E(t_1 + t_2) = \frac{1}{2}[E(t_1) + E(t_2)] = \theta$$

Thus t is also an unbiased estimator of θ. And

$$Var(t) = Var\left[\frac{1}{2}(t_1 + t_2)\right] = \frac{1}{4}Var(t_1 + t_2)$$

$$= \frac{1}{4}[Var(t_1) + Var(t_2) + 2Cov(t_1, t_2)]$$

$$= \frac{1}{4}\left[Var(t_1) + Var(t_2) + 2\rho\sqrt{Var(t_1)Var(t_2)}\right]$$

$$= \frac{1}{2}\ Var(t_1)(1 + \rho),$$

whereρ is Karl Pearson's co-efficient of correlation between $t_1$and t_2 . Since $t_1$is *MUV* estimator, therefore $Var(t_1) \leq Var(t)$. This gives

$$Var(t_1) \leq \frac{1}{2}\ Var(t_1)(1 + \rho)$$

And hence $\rho \geq 1$. Note that if ρ is Karl Pearson's co-efficient, then $|\ \rho\ | \leq 1$. Thus in present case we must have $\rho = 1$. This implies that t_1 and t_2 have linear relation of the form: $t_2 = a + bt_1$, where a and b are constants, independent of sample values $x_1, x_2, \ldots \ldots x_n$ but may depend on parameter θ. Since $t_2 = a + bt_1$, taking expectation on both sides we have, $\theta = a + b\theta$.

And $Var(t_2) = Var(a + bt_1) = b^2 Var(t_1)$

$$\Rightarrow b^2 = 1 \quad \{\because Var(t_1) = Var(t_2)\}$$

$$\Rightarrow \mathrm{b} = \pm 1.$$

But here, $\rho = 1$. Therefore the coefficient of regression of t_1 and t_2 must be positive. Hence b= 1. This implies that a $=$ 0. Consequently, we get $t_1 = t_2$.

14.4.4. SUFFICIENCY:-

An estimator is said to be sufficient for a parameter. if it contains all the information in the sample regarding the parameter. More precisely, if $t = t(x_1, x_2, \ldots \ldots x_n)$ is an estimator of a parameter θ, based on a sample $x_1, x_2, \ldots \ldots x_n$, of size n from the population with p.d.f. $f(x, \theta)$

such that the conditional distribution of $x_1, x_2, \ldots \ldots x_n$ given t, is independent of θ, then t is sufficient estimator for θ.

To understand the concept of sufficiency, consider a random sample $x_1, x_2, \ldots \ldots x_n$ from a Bernoulli population with parameter p, where $0 < p < 1$. Clearly each $x_i{}'$s follows the distribution of population. Therefore, for each $i = 1,2, \ldots., n$

$$x_i = \begin{cases} 1, \text{with probability } p \\ 0, \text{with probability } (1 - p). \end{cases}$$

Define a statistic $t = x_1 + x_2 +, \ldots \ldots + x_n$. Then t follows binomial distribution $B(n, p)$. And hence for $0 \leq k \leq n$,

$$P(t = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Further the conditional distribution of $(x_1, x_2, \ldots \ldots x_n)$ given t is

$$P(x_1 \cap x_2 \cap \ldots \cap x_n | t = k) = \frac{P(x_1 \cap x_2 \cap \ldots \cap x_n \cap t = k)}{P(t = k)}$$

$$= \begin{cases} \dfrac{p^k (1-p)^{n-k}}{\binom{n}{k} p^k (1-p)^{n-k}} = \dfrac{1}{\binom{n}{k}} \\ 0, \quad \text{if} \sum_{i=1}^{n} x_i \neq k \end{cases}$$

Since the conditional distribution of $(x_1, x_2, \ldots \ldots x_n)$ given t, does not depend on parameter p, therefore statistic $t = \sum_{i=1}^{n} x_i$ is sufficient estimator for parameter p.

Theorem14.4.4.1 Neyman Factorization Theorem.:The necessary and sufficient condition for a distribution to admit sufficient statistic is provided by the 'factorization theorem' due to Neyman.

Statement$T = t(x)$ is sufficient for θ if and only if the joint density function L (say), of the sample values can be expressed in the form

$$L = g\theta[t(x)].h(x)$$

where (as indicated) gθ [t(x)] depends on θ and x only through the value of $t(x)$and $h(x)$ is independent of θ.

Invariance Property of Sufficient Estimator.If T is a sufficient estimator for the parameter θ and if φ(T) is a one to one function of T, then φ (T) is sufficient for φ (θ).

Fisher-Neyman Criterion. A statistic $t_1 = t_2(x_1, x_2, \ldots \ldots x_n)$ is sufficient estimator of parameter 0 if and only if the likelihood function (joint p.df. of the sample) can be expressed as:

$$L = \prod_{i=1}^{n} f(x_1, \theta)$$

$$= g_1(t_1, \theta). k(x_1, x_2, \ldots \ldots x_n)$$

where$g_1(t_1, \theta)$ is the $\rho.d.f.$ of statistic t_1 and $k(x_1, x_2, \ldots \ldots x_n)$ is a function of sample observations only independent of θ. Note that this method requires the working out of the p.d.f. (p.m.f.) of the statistic$t_{1=}t(x_1, x_2, \dots \dots x_n)$, which is not always easy.

Example: Let $x_1, x_2, \dots \dots x_n$ be a random sample from $N(\mu, \sigma^2)$ population. Find sufficient estimators for $\mu \; and \; \sigma^2$.

Solution. Let us write

$$\theta = (\mu, \sigma^2); -\infty < \mu < \infty, \qquad 0 < \sigma^2 < \infty$$

Then , L=$\prod_{i=1}^{n} f_\theta(x_i) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n . \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right]$

$=\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n . \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n} {x_i}^2 - 2\mu\sum x_i + n\mu^2\right]$

$=g_\theta[t(x).h(x)]$

Where $g_\theta[t(x)]=\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n . \exp\left[-\frac{1}{2\sigma^2}\{t_1(x) - 2\mu t_2(x) + n\mu^2\right.$

$t(x)=[t_1(x), t_2(x)]=(\sum x_i, \sum {x_i}^2)$ and h(x)=1

Thus $t(x)=\sum x_i$ is sufficient for μ and $t_2(x)=\sum {x_i}^2$, is sufficient for σ^2.

Check your Progress

1. Properties of a good point estimator includes which of the following?
   A. Stationarity
   B. Efficiency
   C. Consistency
   D. Neutrality
   E. Unbiasedness

   Choose the **correct** answer from the options given below:

   i A, B and C only
   ii C, D and E only
   iii A, D and E only
   iv B, C and E only

2. A point estimate is a single value used to estimate a population parameter T\F.

3. A point estimate is a range of values used to estimate a population parameter T\F.

4.An interval estimate is a single value used to estimate a population Parameter T\F.

5.An interval estimate is a range of values used to estimate a population Parameter T\F.

14.5.SOLVED EXAMPLES:-

Example 14.5.1. Consider a random sample $X_1, X_2, \dots, X_n$ from a population which follows Bernoulli distribution with parameter p (i.e. if X is population variate, then X take the value 1 with probability p

and the take the value 0 with probability $(1-p)$. Let $t = \frac{\sum_{i=1}^{n} x_i}{n} \times \left(1 - \frac{\sum_{i=1}^{n} x_i}{n}\right)$. Then t is a consistent estimator of $p(1-p)$.

Solution. Given that $X_1, X_2, \dots, X_n$ are i.i.d Bernoulli r.v. with parameter p. Here define $t_1 = \sum_{i=1}^{n} x_i$. Then t_1 follows binomial distribution $B(n,p)$. It is well known that $E(t_1) = np$and $Var(t_1) = np(1-p)$. Also, in the present case the sample mean is

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{t_1}{n}.$$

This implies that $E(\bar{X}) = \frac{1}{n}\, E\,(t_1) = \frac{1}{n} \times np = p$ and $Var\,(\bar{X}) = Var\,(t_1/n) = \frac{1}{n^2}.Var\,(t_1) = \frac{p(1-p)}{n}$.

From above two equations $E(\bar{X}) \to p$ and $Var(\bar{X}) \to 0$, as $n \to \infty$. Therefore, $\bar{X}$ is a consistent estimator of p. Also we have given that $t = \frac{\sum_{i=1}^{n} x_i}{n} \times \left(1 - \frac{\sum_{i=1}^{n} x_i}{n}\right)$. Then $t = \bar{X}(1-\bar{X})$. Since $\bar{X}$ is consistent estimator of $\bar{p}$, therefore by the invariance property of consistent estimators $\bar{X}(1-\bar{X})$ is a consistent estimator of $p(1-p)$.

Example 14.5.2: Consider a random sample of size 5: X_1, X_2, X_3, X_4, X_5 drawn from a normal population with unknown mean μ. Consider the following estimators for population mean μ:

i. $t_1 = \frac{X_1+X_2+X_3+X_4+X_5}{5}$
ii. $t_2 = \frac{X_1+X_2}{2} + X_3$
iii. $t_3 = \frac{2X_1+X_2+\lambda X_3}{3}$

where λ is such that t_3 is an unbiased estimator of μ. Find the value of λ. Find whether t_1 and t_2 are unbiased or not. Which estimator is best among t_1, t_2 and t_3 for μ, give explanation.

Solution. Let variance of the population is σ^2. We already know that $E(X_i) = \mu, Var(X_i) = \sigma^2$ and $Cov\left(X_i, X_j\right) = 0$, for each $(\,i \neq j = 1,2,\dots,n)$.

Now,$E(t_1) = E\left(\frac{X_1+X_2+X_3+X_4+X_5}{5}\right) = \frac{1}{5}[E(X_1) + E(X_2) + E(X_3) + E(X_4) + E(X_5)] = \frac{1}{5}[\mu + \mu + \mu + \mu + \mu] = \mu$.

Hence $t_1$is an unbiased estimator of μ.Again,

$$E(t_2) = E\left(\frac{X_1+X_2}{2} + X_3\right) = \frac{1}{2}[E(X_1) + E(X_2)] + E(X_3)$$
$$= \frac{1}{2}[\mu + \mu] + \mu = 2\mu.$$

Thus $t_2$is not an unbiased estimator of μ. To find the value of λ for which t_3 is an unbiased estimator $E(t_3) = \mu$.This implies $E\left(\frac{2X_1+X_2+\lambda X_3}{3}\right) = \mu$.Which gives $2E(X_1) + E(X_2) + \lambda E(X_3) = 3\mu$ i.e. $2\mu + \mu + \lambda\mu = 3\mu \Rightarrow \lambda = 0$. Now, consider the $\lambda = 0$. Then

$$Var(t_1) = Var\left(\frac{X_1 + X_2 + X_3 + X_4 + X_5}{5}\right)$$
$$= \frac{1}{25}[Var(X_1) + Var(X_2) + \cdots + Var(X_5)]$$
$$= \frac{1}{25}[\sigma^2 + \sigma^2 + \sigma^2 + \sigma^2 + \sigma^2] = \frac{1}{5}\sigma^2.$$

Similarly, $Var(t_2) = Var\left(\frac{X_1+X_2}{2} + X_3\right)$
$$= \frac{1}{4}Var(X_1 + X_2) + Var(X_3)$$
$$= \frac{1}{4}[Var(X_1) + Var(X_2)] + Var(X_3)$$
$$= \frac{1}{4}[\sigma^2 + \sigma^2] + \sigma^2 = \frac{3}{2}\sigma^2$$ and $Var(t_3) = Var\left(\frac{2X_1+X_2}{3}\right) =$
$$\frac{4}{9}Var(X_1) + \frac{1}{9}Var(X_2)$$
$$= \frac{4}{9}\sigma^2 + \frac{1}{9}\sigma^2 = \frac{5}{9}\sigma^2.$$

After the consideration of λ = 0, only t_1 and t_3 are unbiased estimator, out of which variance of t_1 is lesser. Therefore t_1 is the best estimator (in between t_1, t_2 and t_3) of μ.

Example 14.5.3. Consider a random sample of size 3: X_1, X_2, X_3 from a population with mean value μ and variance σ^2. Let t_1, t_2, t_3 be the estimators for μ defined as:

$$t_1 = X_1 + X_2 - X_3,\ t_2 = 2X_1 + 3X_3 - 4X_2 \text{ and } t_3 = (\lambda X_1 + X_2 + X_3)/3$$

In between t_1, t_2 and t_3, which are unbiased estimators? For what value of λ t_3, is an unbiased estimator for μ. And for this value of λ, check that t_3 is consistent estimator or not. And finally decide that, in between t_1, t_2 and t_3, which is the best estimator for μ.

Solution: As we discussed in previous problem, $X_1 X_2$, X_3 is a random sample from a population with mean μ and variance σ^2, therefore for each $i = 1,2, \dots, n$ $E(X_i) = \mu$, $Var(X_i) = \sigma^2$. And $Cov(X_i, X_j) = 0, (i \neq j = 1,2, \dots, n)$. Now to check unbiasedness of t_1, t_2 and t_3: $E(t_1) = E(X_1) + E(X_2) - E(X_3) = \mu + \mu - \mu = \mu$. This implies that, t_1 is an unbiased estimator of μ. $E(t_2) = 2E(X_1) + 3E(X_3) - 4E(X_2) = 2\mu + 3\mu - 4\mu = \mu$ Hence t_2 is also an unbiased estimator for μ. For the value of λ we are given that $E(t_3) = \mu$ i.e. $\frac{1}{3}[\lambda E(X_1) + E(X_2) + E(X_3)] = \mu$. Which further gives that $\frac{1}{3}[\lambda\mu + \mu + \mu] = \mu \Rightarrow \lambda = 1$. Thus, for $\lambda = 1$ t_3 is an unbiased estimator. Now for $\lambda = 1, t_3 = \frac{X_1+X_2+X_3}{3} = \bar{X}$. i.e. t_3 is sample mean. And by Weak Law of Large Numbers, we already know that sample mean is a consistent estimator of population mean μ. Therefore t_3 is a consistent estimator of μ. Further to check best estimator among t_1, t_2 and t_3 (with $\lambda = 1$): $Var(t_1) = Var(X_1) + Var(X_2) + Var(X_3) = 3\sigma^2$,

$$Var(t_2) = 4Var(X_1) + 9Var(X_2) + 16Var(X_3) = 29\sigma^2,$$

and $Var(t_3) = \frac{1}{9}[Var(X_1) + Var(X_2) + Var(X_3)] = \frac{1}{3}\sigma^2$.
Since $Var(t_3)$ is minimum, therefore t_3 is the best estimator in the sence of minimum variance.

14.6.SUMMARY: -

In this unit, we have studied the basic terminology of sample and population. We have also read the types of sampling. And in the last we have introduced the notion of estimates, statistic and parameter.

14.7.GLOSSARY:-

(i) Point Estimator
(ii) Unbiasedness
(iii) Consistency
(iv) Efficiency
(v) Sufficiency
(vi) Random sample
(vii) Mean
(viii) Variance
(ix) Statistic

14.8.REFERENCES:-

1. S. C. Gupta and V. K. Kapoor: *Fundamentals of mathematical statistics*, Sultan Chand & Sons, 2020.
2. Seymour Lipschutz and John J. Schiller :*Schaum's Outline: Introduction to Probability and Statistics*, McGraw Hill Professional, 2017.
3. J. S. Milton and J. C. Arnold: *Introduction to Probability and Statistics* (4th Edition), Tata McGraw-Hill, 2003.
4. https://www.wikipedia.org.

14.9.SUGGESTED READINGS:-

1. Rohatgi, V. K., &Saleh, A. M. E. (2015). *An introduction to probability and statistics*. John Wiley & Sons.
2. A.M. Goon: *Fundamental of Statistics* (7th Edition), 1998.
3. R.V. Hogg and A.T. Craig: *Introduction to Mathematical Statistics*, MacMillan, 2002.
4. R.V. Hogg, Joseph W. Mc Kean and T. Allen: Craig: *Introduction to Mathematical Statistics* (7th edition), Pearson Education, 2013.

5. Irwin Miller and Marylees Miller John E. Freund: *Mathematical Statistics with Applications* (8th Edition). Pearson. Dorling Kindersley Pvt. Ltd. India, 2014.

14.10 TERMINAL QUESTIONS:-

TQ 14.10.1 If t is an unbiased estimator for θ. show that t^2 is unbiased estimator for θ^2.

TQ 14.10.2 Show that $\frac{\{\sum x_i(\sum x_i - 1)\}}{n(n-1)}$ is an unbiased estimator of θ^2, for sample $x_1, x_{2,}, \ldots \ldots, x_n$ drawn on X which takes the values 1 or 0 with respective probabilities θ and $(1-\theta)$.

TQ 14.10.3 (i). Show that if a most efficient estimator A and a lessefficient estimator B with efficiency e tend to joint normality for large samples, then $B - A$ tends to zero correlation with A.

(ii). Show that the error in B may be regarded as composed (for large samples) of two parts which are independent, the error in A and the error in$(B - A)$.

TQ 14.10.4 Ift_1 *and* t_2 are two unbiased estimators of θ, having the same variance and ρ correlation between them, then show that $\rho \geq 2e - 1$, whare e is the efficiency of each estimator.

TQ 14.10.5 Let $x_1, x_{2,}, \ldots \ldots, x_n$ be a random sample from $N(\mu, \sigma^2)$population, Find sufficient estimators forμ andσ^2.

TQ 14.10.6 Let$x_1, x_{2,}, \ldots \ldots, x_n$ be a random sample from a population with p.d.f.
$f(x, \theta) = \theta x^{\theta-1}$, $0 < x < 1, \theta > 0$. Show that $t = \prod_i^n x_i$ is sufficient for θ.

14.11. ANSWER:-

Answer of check your progress:-

CYQ 1 Option 4.
CYQ 2 True
CYQ 3 False
CYQ 4 False
CYQ 5 True

Answer of Terminal Questions:-

TQ 14.10.5. $\sum x_i$ is sufficient for μ and $\sum {x_i}^2$ is sufficient for σ^2.

Table 1: Normal Table

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998
3.5	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998
3.6	0.4998	0.4998	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.7	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.8	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.9	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000

Table2: t − Distribution table

Significant level (α)

Degrees of freedom (df)	.2	.15	.1	.05	.025	.01	.005	.001
1	3.078	4.165	6.314	12.706	25.452	63.657	127.321	636.619
2	1.886	2.282	2.920	4.303	6.205	9.925	14.089	31.599
3	1.638	1.924	2.353	3.182	4.177	5.841	7.453	12.924
4	1.533	1.778	2.132	2.776	3.495	4.604	5.598	8.610
5	1.476	1.699	2.015	2.571	3.163	4.032	4.773	6.869
6	1.440	1.650	1.943	2.447	2.969	3.707	4.317	5.959
7	1.415	1.617	1.895	2.365	2.841	3.499	4.029	5.408
8	1.397	1.592	1.860	2.306	2.752	3.355	3.833	5.041
9	1.383	1.574	1.833	2.262	2.685	3.250	3.690	4.781
10	1.372	1.559	1.812	2.228	2.634	3.169	3.581	4.587
11	1.363	1.548	1.796	2.201	2.593	3.106	3.497	4.437
12	1.356	1.538	1.782	2.179	2.560	3.055	3.428	4.318
13	1.350	1.530	1.771	2.160	2.533	3.012	3.372	4.221
14	1.345	1.523	1.761	2.145	2.510	2.977	3.326	4.140
15	1.341	1.517	1.753	2.131	2.490	2.947	3.286	4.073
16	1.337	1.512	1.746	2.120	2.473	2.921	3.252	4.015
17	1.333	1.508	1.740	2.110	2.458	2.898	3.222	3.965
18	1.330	1.504	1.734	2.101	2.445	2.878	3.197	3.922
19	1.328	1.500	1.729	2.093	2.433	2.861	3.174	3.883
20	1.325	1.407	1.725	2.086	2.423	2.845	3.153	3.850
21	1.323	1.494	1.721	2.080	2.414	2.831	3.135	3.819
22	1.321	1.492	1.717	2.074	2.405	2.819	3.119	3.792
23	1.319	1.489	1.714	2.069	2.398	2.807	3.104	3.768
24	1.318	1.487	1.711	2.064	2.391	2.797	3.091	3.745
25	1.316	1.485	1.708	2.060	2.385	2.787	3.078	3.725
26	1.315	1.483	1.706	2.056	2.379	2.779	3.067	3.707
27	1.314	1.482	1.703	2.052	2.373	2.771	3.057	3.690
28	1.313	1.480	1.701	2.048	2.368	2.763	3.047	3.674
29	1.311	1.479	1.699	2.045	2.364	2.756	3.038	3.659
30	1.310	1.477	1.697	2.042	2.360	2.750	3.030	3.646
40	1.303	1.468	1.684	2.021	2.329	2.704	2.971	3.551
50	1.299	1.462	1.676	2.009	2.311	2.678	2.937	3.496
60	1.296	1.458	1.671	2.000	2.299	2.660	2.915	3.460
70	1.294	1.456	1.667	1.994	2.291	2.648	2.899	3.435
80	1.292	1.453	1.664	1.990	2.284	2.639	2.887	3.416
100	1.290	1.451	1.660	1.984	2.276	2.626	2.871	3.390
1000	1.282	1.441	1.646	1.962	2.245	2.581	2.813	3.300
Infinite	1.282	1.440	1.645	1.960	2.241	2.576	2.807	3.291

Table 3: F −Distribution Table

F-table of Critical Values of α = 0.01 for F(df1, df2)

	DF1=1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
DF2=1	4052.18	4999.50	5403.35	5624.58	5763.65	5858.99	5928.36	5981.07	6022.47	6055.85	6106.32	6157.29	6208.73	6234.63	6260.65	6286.78	6313.03	6339.39	6365.86
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.49	99.50
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.05	26.87	26.69	26.60	26.51	26.41	26.32	26.22	26.13
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	13.75	10.93	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.67	3.59	3.51	3.43	3.34	3.26	3.17
14	8.86	6.52	5.56	5.04	4.70	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.90	3.81	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.85	2.75
17	8.40	6.11	5.19	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.84	2.75	2.65
18	8.29	6.01	5.09	4.58	4.25	4.02	3.84	3.71	3.60	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	8.19	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.93	2.84	2.76	2.67	2.58	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.86	2.78	2.70	2.61	2.52	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23	7.88	5.66	4.77	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	7.77	5.57	4.68	4.18	3.86	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	2.96	2.82	2.66	2.59	2.50	2.42	2.33	2.23	2.13
27	7.68	5.49	4.60	4.11	3.79	3.56	3.39	3.26	3.15	3.06	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	2.10
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06
29	7.60	5.42	4.54	4.05	3.73	3.50	3.33	3.20	3.09	3.01	2.87	2.73	2.57	2.50	2.41	2.33	2.23	2.14	2.03
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.67	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.81
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.04	1.95	1.86	1.76	1.66	1.53	1.38
∞	6.64	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.19	2.04	1.88	1.79	1.70	1.59	1.47	1.33	1.00

Table 4: CHI-SQUARE Table

Percentage Points of the Chi-Square Distribution

Degrees of Freedom	Probability of a larger value of x^2								
	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
1	0.000	0.004	0.016	0.102	0.455	1.32	2.71	3.84	6.63
2	0.020	0.103	0.211	0.575	1.386	2.77	4.61	5.99	9.21
3	0.115	0.352	0.584	1.212	2.366	4.11	6.25	7.81	11.34
4	0.297	0.711	1.064	1.923	3.357	5.39	7.78	9.49	13.28
5	0.554	1.145	1.610	2.675	4.351	6.63	9.24	11.07	15.09
6	0.872	1.635	2.204	3.455	5.348	7.84	10.64	12.59	16.81
7	1.239	2.167	2.833	4.255	6.346	9.04	12.02	14.07	18.48
8	1.647	2.733	3.490	5.071	7.344	10.22	13.36	15.51	20.09
9	2.088	3.325	4.168	5.899	8.343	11.39	14.68	16.92	21.67
10	2.558	3.940	4.865	6.737	9.342	12.55	15.99	18.31	23.21
11	3.053	4.575	5.578	7.584	10.341	13.70	17.28	19.68	24.72
12	3.571	5.226	6.304	8.438	11.340	14.85	18.55	21.03	26.22
13	4.107	5.892	7.042	9.299	12.340	15.98	19.81	22.36	27.69
14	4.660	6.571	7.790	10.165	13.339	17.12	21.06	23.68	29.14
15	5.229	7.261	8.547	11.037	14.339	18.25	22.31	25.00	30.58
16	5.812	7.962	9.312	11.912	15.338	19.37	23.54	26.30	32.00
17	6.408	8.672	10.085	12.792	16.338	20.49	24.77	27.59	33.41
18	7.015	9.390	10.865	13.675	17.338	21.60	25.99	28.87	34.80
19	7.633	10.117	11.651	14.562	18.338	22.72	27.20	30.14	36.19
20	8.260	10.851	12.443	15.452	19.337	23.83	28.41	31.41	37.57
22	9.542	12.338	14.041	17.240	21.337	26.04	30.81	33.92	40.29
24	10.856	13.848	15.659	19.037	23.337	28.24	33.20	36.42	42.98
26	12.198	15.379	17.292	20.843	25.336	30.43	35.56	38.89	45.64
28	13.565	16.928	18.939	22.657	27.336	32.62	37.92	41.34	48.28
30	14.953	18.493	20.599	24.478	29.336	34.80	40.26	43.77	50.89
40	22.164	26.509	29.051	33.660	39.335	45.62	51.80	55.76	63.69
50	27.707	34.764	37.689	42.942	49.335	56.33	63.17	67.50	76.15
60	37.485	43.188	46.459	52.294	59.335	66.98	74.40	79.08	88.38

Teen Pani Bypass Road, Transport Nagar
Uttarakhand Open University, Haldwani, Nainital-263139
Phone No. 05946-261122, 261123
Toll free No. 18001804025
Fax No. 05946-264232,
E-mail:info@uou.ac.in
Website: https://www.uou.ac.in/