PROBABILITY THEORY AND DISTRIBUTIONS

M.Sc., STATISTICS First Year

Semester – I, Paper-I

Lesson Writers

Prof. G. V. S. R. Anjaneyulu
Professor of Statistics(Retired)
Department of Statistics
University College of Sciences
Acharya Nagarjuna University

Dr. R. Vishnu Vardhan
Associate Professor
Department of Statistics
Pondicherry University

Dr. Syed Jilani
Department of Statistic
University College of Sciences
Acharya Nagarjuna University

Dr. Bala Naga Hima Bindu. I
Sr. Asst. Professor of Statistics
Dept. of Humanities & Sciences,
CVR College of Engineering,
Hyderabad,

.

Editor:
Dr. R. Vishnu Vardhan
Associate Professor
Department of Statistics
Pondicherry University

Director, I/c
Prof. V.VENKATESWARLU
MA., M.P.S., M.S.W., M.Phil., Ph.D.

CENTRE FOR DISTANCE EDUCATION
ACHARAYANAGARJUNAUNIVERSITY
NAGARJUNANAGAR – 522510
Ph:0863-2346222,2346208,
0863-2346259(Study Material)
Website: www.anucde.info
e-mail:anucdedirector@gmail.com

M.Sc., STATISTICS - PROBABILITY THEORY AND DISTRIBUTIONS

First Edition 2025

No. of Copies :

This book is exclusively prepared for the use of students of M.SC.(Statistics) Centre for Distance Education, Acharya Nagarjuna University and this book is meant for limited Circulation only.

Published by:

Prof. V.VENKATESWARLU,

Director I/C

Centre for Distance Education,

Acharya Nagarjuna University

Printed at:

FOREWORD

Since its establishment in 1976, Acharya Nagarjuna University has been forging ahead in the path of progress and dynamism, offering a variety of courses and research contributions. I am extremely happy that by gaining 'A^+' grade from the NAAC in the year 2024, Acharya Nagarjuna University is offering educational opportunities at the UG, PG levels apart from research degrees to students from over 221 affiliated colleges spread over the two districts of Guntur and Prakasam.

The University has also started the Centre for Distance Education in 2003-04 with the aim of taking higher education to the doorstep of all the sectors of the society. The centre will be a great help to those who cannot join in colleges, those who cannot afford the exorbitant fees as regular students, and even to housewives desirous of pursuing higher studies. Acharya Nagarjuna University has started offering B.Sc., B.A., B.B.A., and B.Com courses at the Degree level and M.A., M.Com., M.Sc., M.B.A., and L.L.M., courses at the PG level from the academic year 2003-2004 onwards.

To facilitate easier understanding by students studying through the distance mode, these self-instruction materials have been prepared by eminent and experienced teachers. The lessons have been drafted with great care and expertise in the stipulated time by these teachers. Constructive ideas and scholarly suggestions are welcome from students and teachers involved respectively. Such ideas will be incorporated for the greater efficacy of this distance mode of education. For clarification of doubts and feedback, weekly classes and contact classes will be arranged at the UG and PG levels respectively.

It is my aim that students getting higher education through the Centre for Distance Education should improve their qualification, have better employment opportunities and in turn be part of country's progress. It is my fond desire that in the years to come, the Centre for Distance Education will go from strength to strength in the form of new courses and by catering to larger number of people. My congratulations to all the Directors, Academic Coordinators, Editors and Lesson-writers of the Centre who have helped in these endeavors.

Prof. K.GangadharaRao
M.Tech.,Ph.D.,
Vice-Chancellor I/c
Acharya Nagarjuna University

M.Sc. – Statistics Syllabus
SEMESTER-I
101ST24: Probability Theory and Distributions

Unit-I: Classes of sets, fields, minimal σ fields, sequence of sets, limit supremum and limit infimum of sequence of sets, measure, probability measure, properties of measure, axiomatic definition of probability, continuity theorem of probability, conditional probability, statistical independence of events, probability on finite sample spaces, geometrical probability.

Unit-II: Measurable functions, notation of random variable, distribution function, properties of distribution, vector of random variables, statistical independence, concepts of joint, marginal and conditional distributions, mathematical expectation, conditional expectation, characteristic function, its properties. Inversion formula, characteristic functions and moments. Moments inequalities-Markov, Schwartz, Jenson, Holder, Minkowski, Kolmogrove's, Hajek-Renyi.

Unit-III: Convergence of sequence of random variable- Type of convergence-in probability, almost sure, in mean square, in law- their interrelations. Law of large numbers-weak laes: chebychevs's form of W.L.L.N., Necessary and Sufficient Condition of W.L.L.N. Kinchines form of W.L.L.N., Kolmogrove's S.L.L.N. for i.i.d. Random Variables.

Unit-IV: Discrete distributions – Compound Binomial, Compound poisson, multinomial, truncated Binomial, truncated poisson distributions and their properties. Continuous distributions – Laplace, Weibull, Logistic and Pareto distributions and their properties.

Unit-V: Order statistics- distribution function, probability density function (p.d.f.) of single order statistic, joint p.d.f. of order statistics. Distribution of range with applications in rectangular and exponential cases.

BOOKS FOR STUDY:

1) Modern probability theory by B.R.bhat, Wiley Eastern Limited.
2) An introduction to probability theory and mathematical statistics by V.K.Rohatgi, John Wiley.
3) An Outline of statistics theory-1, by A.M.GOON, M.K.Gupta and B.Dasgupta, the world Press Private Limited, Calcutta.
4) The Theory of Pronanility by B.V.Gnedenko, MIR Publishers, Moscow.
5) Discrete distributions – N.L.Johnson and S.Kotz, John Wiley & Sons.
6) Continuous Univariate distributions, vol. 1 & 2- N.L.Johnson and S.Kotz, John Wiley & Sons.
7) Mathematical Statistics-Parimal Mukopadhyay, New Central Book Agency (P) ltd., Calcutta.

REFERENCE BOOKS:

1) Billingsley,P. (1986): Probability and Measure. Wiley.
2) Kingman,J F C and Taylor, S.J. (1966): Introduction to Measure and Probability. Cambridge University Press.
3) David, H.A (1981): order Statistics, 2nd Ed, John Wiley.
4) David, H.A and Nagaraja H.N. (2003): Order Statistics, 3/e, john Wiley & Sons.
5) Feller, W (1966): Introduction to probability theory and its applications, Vol.II, Wiley.
6) Cramer H (1946); Mathematical Methods of Statistics, Princeton.

CODE: 101ST24

M.Sc DEGREE EXAMINATION

First Semester

Statistics::Paper I – PROBABILITY THEORY AND DISTRIBUTIONS

MODEL QUESTION PAPER

Time : Three hours **Maximum : 70 marks**

Answer ONE question from each Unit. **(5 x 14 = 70)**

UNIT – I

1. (a)(i) Given a class $\{A_i, i=1,2,...,n\}$ of n sets prove that the re-exists a class $\{B_i, i=1,2,...,n\}$ of n disjoint sets such that their unions are equal.
   (ii) Give the axiomatic definition of probability. State the various properties of probability.
   (b) Define probability measure. Establish its simple properties.

OR

2. (a) Write short notes on (i) σ fields (ii) conditional probability (iii) discrete probability space and give one example in each case.
   (b) State and prove Borel- Cantellilemma.

UNIT – II

3.(a) Define characteristic function. State and prove inversion formula.
   (b) State and prove Holder's inequality. Hence, obtain Schwartz inequality.

OR

4. (a) Define distribution function. State and prove its properties.
   (b) State and prove i) Jensen and ii) Markov inequalities.

UNIT – III

5. (a) State and prove Chebychev's form of weak law of large numbers.
   (b) Explain the types of convergence. Prove that almost sure convergence implies convergence in probability.

OR

6. (a) State and prove Kolmogorov's strong law of large numbers.
   (b) State and prove Kinchine's form of W.L.L.N.

UNIT – IV

7. (a) Derive the distribution of compound binomial.
   (b) Obtain the m.g.f. of truncated Poisson distribution and hence, find its mean and variance.

OR

8. (a) Define Weibull distribution. Find its m.g.f., mean and variance.
   (b) Define Laplace distribution. Obtain its characteristic function, mean and variance

UNIT – V

9. (a) Derive the distribution of range.
   (b) Explain p.d.f of single order statistics and Joint order statistics.

OR

10. (a) Let $x_1, x_2, \ldots x_n$ be the set of order statistics of independent random variables, $x_1, x_2, \ldots\ldots x_n$ With common p.d.f. $f(x) = \beta e^{-x}\beta, x \geq 0$.
    (b) Write a the application of distribution of range in rectangular and exponential cases.

CONTENTS

LESSON -1

SETS AND CLASSES OF EVENTS

OBJECTIVES:

After studying this unit, you should be able to:

- To understanding the sets and classes of events
- To know the concept of Structure and sets and classes of events
- To acquire knowledge about significance of sets and classes of events
- To understand the purpose and objectives of pivotal provisions of the sets and classes of events

STRUCTURE:

1.1. INTRODUCTION:

In everyday life we come across many phenomena, the nature of which cannot be predicted in advance, or many experiments, whose outcomes may not be known precisely. However we may know that the outcome has to be one of the several possibilities. The weight of a newborn baby cannot be known before the birth, except that nay lie in a certain range. When a coin is tossed, we known that the outcome has to be either a head or a tail; but we do not know the outcome of a particular throw in advance.

By a statistical experiment or simply experiment we mean not only an experiment such as the tossing of coin or the observation of the number of defects in a certain sample of N items chosen from the daily production, in which the possible outcomes are finite; but also the observation of a phenomenon such as the weight of a new born baby, or the weather condition of a certain region, where the number of possible outcomes is infinite. The concept of experiment is fundamental to the study of probability theory, because it is concerned with assigning chance to the outcomes of the statistical experiment or to possible state of nature and studying them.

Let us denote by ω, the typical outcome of an experiment E; ω is called a sample point. The totally of all outcomes of E will be denoted by **Ω** and is called the sample space.

1.2 EVENT:

A collection of outcomes of a statistical experiment **E** in which we are interested is called an event.

Thus an event is a subset of **Ω**. The number of sample points may be finite, countable or uncountable as illustrated by the following examples.

Example.1: Thus in throwing a die, the results-face 5 turns up, an even number turns up are both events.

Events may be of two types:

1. Elementary Event (or) case: It is an event which cannot be broken further into smaller event .it is an "atom" of event. Thus, in throwing a die, each of the events 1,……..6 is an elementary event. Symbolically, an elementary event will be often be represented by case letters a, b, etc.

2. Random(contingent) Event: A random event is obtained through the combination of several elementary events. Thus, in throwing a die, the event that the odd number turns up is combination of three elementary events:1 or 3 or 5 turns up. This event will be often be represented by large case letters A, B, etc

We shall often use simply the word event. Whether it is elementary or compound would be clear the context.

3. Mutually Exclusive Events: Events is said be mutually exclusive (or incompatible) with respect to an experiment if the occurrence of one of them precludes the occurrence of all the other every time the experiment is performed; in other words the two events cannot be materializing simultaneously. Thus throwing a die, the events, 2 occurs and 5 occurs are mutually exclusive.

4. Exhaustive Events: A set of events in relation to be random experiment is said to be exhaustive if one of them must necessarily materialize every time the experiment is performed. Thus in throwing a die case 1,…..,6 from on exhaustive set of cases.

5. Equally likely Cases: cases are said to be equally likely when we have no reason to believe that one is more likely occur than the other. Thus in drawing a card at random from a full pack of well-shuffled cards, one may believe that each of the cards is equally likely to appear and in case, each of the 52 cases is equally likely. This concept, therefore, presupposes the simplest hypothesis regarding the possibility of occurrences of different cases viz. each case is equally likely to materialize.

6. Favourable Cases: A case **'a'** is said to be favourable to an event A if whenever 'a' occurs. Thus in throwing a die, each of this cases 1,3,5 is favourable to the event, 'an odd face turns up' .

Example .2: a head and a tail being denoted by H and T respectively, the cases HH, HT, TH, TT are mutually exclusive, exhaustive and equally likely cases for the random experiment of throwing a fair coin twice.

Example.3: In throwing two balanced dice, mutually exclusive, exhaustive and equally likely cases are 36 in number namely $(1,1),(1,2),.....,(1,6),(2,1),.....,(2,6),.....,(6,1),....,(6,6)$, where in $(x,y),x,$ y denote the result of first and second throw respectively.

1.3 ALGEBRA OF SETS:

Set: A set is a collection of some elements which are it members. Thus we may have a set of college students, a set of some books, a set of some digits, a set of houses etc. A set can be denoted either by enumerating all its elements given in a brace or by indicating (also in braces) a rule its associates an element with a set. Thus the set of all odd numbers not exceeding 15 can be written as $\{1,3,5,7,9,11,13,15\}$ or $\{2x+1; x=1,2,...,7\}$. A set is often pictured as an oval shaped space with points denoting its elements. The diagram representing a set is called a Venn diagram.

Depending on how many elements it has, a set may be finite or infinite. The set $M=\{1,2,...,50\}$ is finite and contains 50 elements. The set of all natural numbers $N_1=\{1,2,...,n,...\}$ is infinite. The set of all even numbers $N_2=\{2,4,...,2n,...\}$ is also infinite. An infinite set is said to be countable if all its elements can be enumerated. Both the sets N_1,N_2 above are countable. The set C is all points within or on a circle of radius $r>0, C=\{(x,y): x^2+y^2\leq r^2\}$ is infinite and uncountable. Its elements cannot be enumerated.

Universal set: It is the largest possible set of interested under a given condition. All other set are subsets of the universal sets. Thus if we are interested in some students of a college, a set containing all the students of this college may constitute an universal set. The set of all first year commerce students, set of all second year statistics students are all subsets of the above universal set. An universal set is often represented as Ω.

Null set: It is a set containing no element. It is also called an empty set and often represented as ϕ. It is a member of all other sets.

Disjoint sets: Two sets A ,B are disjoint if they have no common element. Similarly sets $A_1,...,A_m$ are mutually disjoint if no two of them have any common element.

In the probability theory an elementary event or a case resulting from a random experiment is represented as a point in a set. A compound event A is represented as a set containing all the points which stand for the elementary events of which A is made up(i.e. which are favorable to A). Thus in throwing a die, each of the cases 1,…,6 is a point in the set. The compound event, an odd face turns up is representing them are disjoint. Events represented by the sets A,B,C are mutually exclusive.

The set containing all the possible points representing the elementary events of a random experiment i.e. the universal set is called the ‘sample space’. It is often represented as S. Thus in tossing a coin once, $S=\{H,T\}$. In tossing a coin twice, $S=\{HH,HT,TH,TT\}$. In throwing two dice, $S=\{(1,1),(1,2),...,(6,6)\}$ represents the sample space for the throw of two dice, set B denoting the event that the total of the two faces is 9.

Example: A coin is tossed until a head appears. Here the sample space consists of elementary events $H, HT, TTH, TTTH, \dots$. These points are countable and infinite in number. The sample space consists of countably infinite number of cases.

(a) Set-operations

Since events are sets of sample points it is essential that one becomes familiar with the algebra of sets, in order to understand manipulations involving events. In the following we assume that Ω is given. The important set operations are:

(i) Complementation;
(ii) Inclusion and Equality;
(iii) Union and Intersection.

(i) Complementation: To every set A we can associate another set A^c consisting of all points of Ω not contained in A. The set A^c is called the complement of A. It denotes the event "*A* does not occur". Symbolically, $A^c = \{\omega \in \Omega / \omega \notin A\}$

Evidently, $\Omega^c = \phi$ and $\phi^c = \Omega$.

Example 1: Suppose a coin is tossed 3 times. The possible outcomes may be denoted by

HHH, HHT, HTH, THH, HTT, THT, TTH, TTT

$\omega 1$, $\omega 2$, $\omega 3$, $\omega 4$, $\omega 5$, $\omega 6$, $\omega 7$, $\omega 8$

If A is the event that at least one head turns up, $A = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6, \omega_7\}$

Then $A^c = \{\omega_8\}$ = {TTT} represents the event that no head turns up.

If B is the event that exactly one head turns up, then

$$B = \{\omega_5, \omega_6, \omega_7\},\ B^c = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_8\}$$

latter represent the event that the number of heads turning up is not equal to one. In general, complements of events are also events. It may be noted that

$$\left(A^c\right)^c = \{\omega \in \Omega / \omega \notin A^c\} = \{\omega \in \Omega / \omega \in A\} = A$$

(ii) a. Inclusion: If all points of a set *A* are also points of another set *B* , then we say that *A* is a subset of *B* or, *A* is included in or contained in *B.* This is denoted by $A \subset B$ or $B \supset A$. Symbolically,

$$A \subset B \Leftrightarrow (\omega \in A \Rightarrow \omega \in B).$$

Evidently

(i) A ⊂ A, (reflexivity);

(ii) A ⊂ B, B ⊂ C ⇒ (ω ∈ A , ω ∈ B ⇒ ω ∈ C) ⇒ A ⊂ C (Transivity).

Example 2:.
If Ω represents the collection of all students, and A is set of all candidates who passed in first class and B is a set of all candidates who passed, then A⊂B. Students who do not pass form a subset of those who do not pass in first class. In general, if A⊂B, then $A^c \supset B^c$

We shall came to know that later that a subset of an event may not be an event .

(ii) b. Equality: If A ⊂ B and B ⊂ A, then A and B are said to be equal, denoted by A=B. Thus in any problem , in order to establish the equality of two sets A and B we have to prove that

$$\omega \in A \Leftrightarrow \omega \in B$$

Since inclusion relation is reflexive and transitive, equality relation is also reflexive and transitive. It is also symmetric, i.e, A = B ⇔ B = A.

(iii) Union and Intersection: If A and B are two sets, then the set of all points ω which belong to either A or B is called A union B and is denoted by A ∪ B. The set of all points which belong to both A and B is called "A intersection B " and is denoted by A ∩ B.

Similarly the union of m sets $A_1, A_2, \ldots, A_m$, namely, $\bigcup_1^m A_i$ is the set of all elements contained in at least one of $A_1, A_2, \ldots, A_m$.

Similarly the intersection of m sets $A_1, A_2, \ldots, A_m$, namely, $\bigcap_{i=1}^m A_i$ or simply $A_1, A_2, \ldots, A_m$ is the set of all elements contained in all of $A_1, A_2, \ldots, A_m$.

Symbolically

A ∪ B = {ω : ω ∈ A or ω ∈ B} = {ω : ω belongs to at least one of the sets A or B}

A ∩ B = {ω : ω ∈ A and ω ∈ B} = {ω : ω ∈ both A and B }.

If A ⊂ B, A ∪ B = B, A ∩ B = A

Example3:
A = {all the 3 tosses result in the same outcome},

= { ω1, ω8 }, and

B = {the 3 tosses have at most one tail },

= { ω1 , ω2 , ω3 , ω4 },

then A ∪ B = { ω1 , ω2 , ω3 , ω4 , ω8 },

A ∩ B = { ω1 }.

If A ∩ B = ∅ then A and B are said to be disjoint or mutually exclusive. In this case and only in this case and only in this case A ∪ B will be denoted by A + B. Thus, A + A^c = Ω. We use this notation throughout the book.

If A ⊂ B, B ∩ A^c will be denoted by B − A and is called the proper difference of B and A. Evidently, B−A and A are disjoint and (B−A) + A = B. We may note that

A^c = Ω − A. Many times A ∩ B is written as AB omitting ∩.

We may note that

AB ⊂ A, A− AB = AB^c ⊂ B^c ;

AB ⊂ B, B− AB = BA^c ⊂ B.

Hence AB^c and BA^c are disjoint. Their union AB^c + BA^c = A Δ B is called the symmetric difference of A and B. Here Δ is called the symmetric difference operation. While the proper difference is defined only if A ⊂ B, symmetric difference is always defined.

We may note that AB + AB^c = A.

Example 4:

Suppose Ω is the real line R consisting of all real points ω,

Ω = { ω : −∞ < ω < ∞}.

Define

A = {ω : ω ∈ (−∞, a)} = {ω < a}, B = {ω : ω ∈ (c, d)} = {c < ω < d},

Then

A ∩B = ∅, if a < c < d,

= (c, a), if c < a < d,

= (c, d), if c < d < a.

Since A ∪B = {ω : either ω < a or c <ω < d}, it will note be an interval if a < c < d, even though A and B are intervals.

1.4 CLASS OF SETS:

A group of sets is termed as class of sets and it is denoted by □ . An example is □ $=\{\{H\},\{T\},\{H,T\}\}$.

LEMMA: Given a class $\{A_i, i=1,2,\ldots,n\}$ of n sets there exists a class $\{B_i, i=1,2,\ldots,n\}$ of disjoint sets such that, $\bigcup_{i=1}^{n} A_i = \sum_{i=1}^{n} B_i$

Proof : The lemma will be proved by induction. Evidently $A_1 \cup A_2 = A_1 + A_1^c A_2 = B_1 + B_2$, (say)

Where B_1 and B_2 are disjoint.

Thus the lemma is true for $n = 2$. Suppose it is true for all $n \le m \le 2$. Then

$$\bigcup_{i=1}^{m+1} A_i = \left(\bigcup_{i=1}^{m} A_i \right) \cup A_{m+1}$$

$$= \left(\sum_{i=1}^{m} B_i \right) \cup A_{m+1} \text{ (By Induction hypothesis)}$$

$$= \sum_{i=1}^{m} B_i + \left(\sum B_i \right)^c A_{m+1}$$

$$= \sum_{i=1}^{m} B_i + B_{m+1}, \; say$$

Where B_{m+1} and $\sum B_i$ are disjoint and hence B_{m+1} and B_i are disjoint for $\mathrm{i} = 1,2,\ldots,\mathrm{m}$. Hence the lemma holds for $\mathrm{n} = \mathrm{m}+1$ and by induction for arbitrary *n*.

Note that $B_i \subset A_i \forall i$

Corollary: $\bigcup_{i=1}^{\infty} A_i = A_1 + A_1^c A_2 + A_1^c A_2^c A_3 + \ldots$

This is the extension of lemma to the countable class. It tells us how to get a countable class of disjoint sets, starting from a countable class of arbitrary sets, such that their unions are equal. Thus, in future, we may assume that we are given a countable class of disjoint sets, without loss of generality

Proof: suppose $\omega \in \bigcup_{i=1}^{\infty} A_i$. Then ω belongs to some A_i. Thus ω may belong to A_1 or ω may belong to A_1^c. In the latter case ω has to belong to A_2 or A_2^c. Thus either $\omega \in A_1^c A_2$ or $\omega \in A_1^c A_2^c$. In the latter case ω has to belong to A_3 or A_3^c. Continuing in this matter, ω has to belong to either $A_1, A_1^c A_2, A_1^c A_2^c A_3, \ldots, A_1^c A_2^c \ldots A_{n-1}^c A_n$ etc. Hence $\omega \in \mathrm{RHS}$ of (corollary)

Conversely, if $\omega \in \mathrm{RHS}$ of $\omega \in A_1^c A_2^c \ldots A_{k-1}^c A_k$ for some k. But $A_1^c A_2^c \ldots A_{k-1}^c A_k \subset A_k$. Hence $\omega \in A_k$ for some k. Thus $\omega \in \bigcup_{i=1}^{\infty} A_k$, which establishes the equivalence of the two sets on the two sides of (corollary)

1.5 FIELD AND MINIMAL FIELD:

Field-Def:

A non empty class of sets □ which is closed under finite unions and complementation is called a field (or a algebra). Thus

$\mathrm{A} \in □, \mathrm{B} \in □ \Rightarrow \mathrm{A} \cup \mathrm{B} \in □$ and $A \in □ \Rightarrow A^c \in □$

THEOREM: A FIELD □ IS CLOSED UNDER FINITE UNIONS. CONVERSLY A CLASS □ CLOSED UNDER COMPLIMENTATION AND FINITE UNIONS IS A FIELD.

PROOF: suppose A is a field then

(i) $A \in \square \Rightarrow A^c \in \square$

(ii) $A_1, A_2, \ldots, A_n \in \square \Rightarrow \bigcap_{i=1}^{n} A_i \in \square$.

But, $A_1, A_2, \ldots, A_n \in \square \Rightarrow A_1^c, A_2^c, \ldots, A_n^c \Rightarrow \square \Rightarrow \bigcap_{i=1}^{n} A_i^c \in \square$, $\Rightarrow \left(\bigcap_{i=1}^{n} A_i^c\right)^c \in \square$, $\Rightarrow \bigcup_{i=1}^{n} A_i \in \square$

Hence □ is closed under finite unions.

(iii) $A_1, A_2, \ldots, A_n \in \square \Rightarrow \bigcup_{i=1}^{n} A_i \in \square$

But, $A_1, A_2, \ldots, A_n \in \square \Rightarrow A_1^c, A_2^c, \ldots, A_n^c \Rightarrow \square \Rightarrow \bigcup_{i=1}^{n} A_i^c \in \square$, $\Rightarrow \left(\bigcup_{i=1}^{n} A_i^c\right)^c \in \square$, $\Rightarrow \bigcap_{i=1}^{n} A_i \in \square$

Hence □ is closed under finite intersections also.

Thus, from this theorem, a field is sometimes defined as a class closed under complementation, finite intersections and/or finite unions.

Evidently, if □ is a field.

$A \in \square \Rightarrow A^c \in \square, \Rightarrow A \bigcup A^c \in \square \Rightarrow A \bigcap A^c \in \square$

Hence, we have the following corollary.

Note: Thus, from the above lemma, we may define a non empty class of sets □ as a field, if □ is closed under complemenations, finite intersections and/or finite unions.

Corollary : Every field □ contains the empty set ϕ and the whole sample space Ω.

PROOF: The class containing only ϕ and Ω is a field. It is the smallest field and is contained in every other field. It is called the degenerate or trivial field and is the power set consisting of every subset of a finite Ω is also a field and is the largest field. If a field contains A, it has to contain A^c and hence contain the class $\{A, A^c, \phi, \Omega\}$.

But this class is a field, contained every field containing A. Therefore it is the smallest field containing A.

Minimal field: consider an arbitrary class S of sets and the smallest field containing S is called the minimal field containing S (or) the field generated by S.

Definition: Let the sequence $\{A_i\}, i = 1, 2, \ldots., n$ be the mutually exclusive and exhaustive then $\{A_i\}, i = 1, 2, \ldots., n$ is called a partition of Ω.

In the case, when $\{A_i\}, i = 1, 2, \ldots., n$ is a partition.

$A_1^c = A_2 + A_3 + \ldots + A_n, (A_1 + A_2)^c = A_1 + A_3 + \ldots + A_n$, and so on.

Hence the class

$$S=\{\phi, A_1, A_2, \ldots, A_n, A_1+A_2, A_1+A_3, \ldots, A_{n-1}+A_n, A_1+A_2+A_3, \ldots, A_{n-2}+A_{n-1}+A_n, \ldots, A_1+A_2+A_3+\ldots+A_n=\Omega\},$$

Containing ϕ, Ω and the unions of A_i's taken one set at a time, 2sets at a time, etc., Which is therefore closed under finite unions, is also closed under complementation. Thus, it is a field. It is the minimal field $F\{A_i\}$ containing $\{A_i\}, i=1,2,\ldots,n$.

It contains

$$\binom{n}{0}+\binom{n}{1}+\binom{n}{2}+---+\binom{n}{n}=\sum_{i=0}^{n}\binom{n}{i}=2^n$$

Sets. This field is also a minimal field containing $\{A_i,\ i=1,2,\ldots,n\}$

We have see that it is very easy to obtain $F(\square)$ the minimal field containing a class $\square$ if ℓ is a partition. In general, to obtain $F(\square)$ we may proceed along the following steps.

(i) obtain $\square_i=\{\varphi,\Omega,A,A^c\}$ such that either $A\in\square\, A(\text{or})A^c\in\square\}$ where $A\subset\Omega$ evidently ℓ is closed under complementation and contain ℓ.

(ii) obtain the class $\square_2$ containing $\bigcap_{k=1}^{n} B_k$, where $B_k\in\square_i, k=1,2,\ldots,n$ B_k's and n being arbitrary now $\square_2$ is closed under finite intersection but not under complementation.

(iii) obtain $\square_3$, the class of all finite unions of pair-wise disjoint subsets belonging to $\square_2$ since they also contain complements $\square_3$ is a field and is the minimal field containing $\square$ (provel)

Example: let$\square=\{\mathrm{A},\mathrm{B}\}$ then

$$\square_1=\{\phi,\Omega,A,A^c,B,B^c\},$$

$$\square_2=\{\square_1, AB, AB^C, A^CB, A^CB^C\},$$

$$\square_3=\{\square_2, AB+A^CB^C, AB^C+A^CB, AB+AB^C+A^CB, AB+AB^C+A^CB^C, AB+A^CB+A^CB^C, AB^C+A^CB+A^CB^C\},$$

$\square_3$ is the smallest field containing $\square$. It coincides with the minimal field containing the partition $\{AB, AB^C, A^CB, A^CB^C\}$. The minimal field containing $\{A,B,C\}$ is the minimal field containing the partition, $\{ABC, A^CBC, AB^CC, ABC^C, A^CB^CC, AB^CC^C, A^CBC^C, A^CB^CC^C\}$.

1.6 $\sigma-$field AND MINIMAL $\sigma-$field:

$\sigma-$field **:** A nonempty class of sets □ which is closed under complementations, and countable unions (or countable intersections) is called a $\sigma-$field and σ _algebra.

Evidently, it is a field closed under countable operations of union and intersection. It possesses all the properties of a field, hence contains null set φ and universal set Ω. **If the class □ contains only a finite number of sets and is a field, it is also a $\sigma-$field. However, a field containing an infinite number of sets may not be a $\sigma-$field.**

Theorem: The intersection an arbitrary number of $\sigma-$fields is a $\sigma-$field. **However, the union of two $\sigma-$fields is not a $\sigma-$field.**

Example: Let $\Omega=\{1,2,3,4\}$ and □ be the class of subsets of A of Ω such that either A contains a finite number of points or A^c contains a finite number of points, evidently □ is closed under complementation. It is also closed under finite unions because $A\cup B$ will contain a finite number of points if each one of A and B is finite and $(A\cup B)^c=A^c\cap B^c$ will contain a finite number of points, if either A^c is finite or B^c is finite. Thus either $A\cup B$ contains a finite number of points or $(A\cup B)^c$ contains a finite number of points. Hence $A\cup B\in$ □ . Thus □ is a field.

But □ is not a σ field.

Let $A_i=\{2i\}(i=1,2,...)$. Then $\bigcup_{i=1}^{\infty}A_i=\{2,4,6...\}$, is neither finite nor the complement is finite,

Hence if does not belong to ξ.

(c) σ-Field or Borel Field:

Closure under finite operators does not implies closure operator, In ex-1.11 we have seen that the class □ of all intervals of from (x,∞), $x\in R$ is closed under finite intersection. But it is not closed under countable intersection, because $\bigcap_{n=1}^{\infty}(n-1)(n\,\infty)=[x,0)\in$ □

A has empty class of sets which is closed under complementation and countable unions (or) uncountable. Intersection is called a σ field evidently if is a field closed under countable operation of union and intersection. It possesses only a finite number of sets and is a field, it is also a σ field. However a field, it is also a σ field containing a infinite number of sets may not be a σ field.

This is the extension of lemma to the countable class. It tells us how to get a countable class of disjoint sets, starting from a countable class of arbitrary sets, such that their unions are equal. Thus, in future, we may assume that we are given a countable class of disjoint sets, without loss of generality

Proof: suppose $w \in \bigcup_{i=1}^{\infty} A_i$. Then w belongs to some A_i. Thus to A_1 or w may belong to A_1^c. In the latter case w has to belong to A_2 or A_2^c. Thus either $w \in A_1^c A_2$ or $w \in A_1^c A_2^c$. In the latter case w has to belong to A_3 or A_3^c. Continuing in this matter, w has to belong to either $A_1, A_1^c A_2, A_1^c A_2^c A_3, \ldots, A_1^c A_2^c \ldots A_{n-1}^c A_n$ etc. Hence $w \in$ RHS of (corollary)

Conversely, if $w \in$ RHS of $w \in A_1^c A_2^c \ldots A_{k-1}^c A_k$ for some k. But $A_1^c A_2^c \ldots A_{k-1}^c A_k \subset A_k$. Hence $w \in A_k$ for some k. Thus $w \in \bigcup_{i=1}^{\infty} A_k$, which establishes the equivalence of the two sets on the two sides of (corollary).

Minimal $\sigma-$ **field:**

An arbitrary class of □ of sets. The smallest $\sigma-$ field containing □ is called minimal $\sigma-$ field

The minimal $\sigma-$ field containing the class □ will be denoted by $\sigma($□$)$. It is the intersection of all $\sigma-$ field containing □. It is also called the $\sigma-$ field generated by □. If □ is finite, the minimal field $F($□$)$ containing □ coincides with the minimal $\sigma-$ field $\sigma($□$)$ containing □.

Consider the class □ of all intervals of the form $(-\infty, x), x \in R$, as subsets of the real line R. This class is closed under finite intersections, but not under complementation nor under countable intersections. Let $\sigma($□$) = B$ be the minimal $\sigma-$ field containing □. Then B contains of the form $[x, \infty)$; which are complements of sets of the form $(-\infty, x)$. It contains intervals of the form $(-\infty, a] = \bigcap_{n=1}^{\infty} \left(-\infty, a + \frac{1}{n}\right)$ (by countable intersection)

$(a, \infty) = (-\infty, a]^c$, (by complementation)

$(a,b) = (-\infty, b) \cap (a, \infty)$ $(a < b)$, $(a,b]$, $[a,b)$, for $a,b \in R$

The $\sigma-$ field $\sigma($□$)$, the minimal $\sigma-$ field containing one of the classes in (1) is called the borel field B of subsets of the real line.

1.7 SEQUENCE OF SETS::

An arrangement of sets in accordance with the set of natural numbers is known as a sequence of sets. More specifically, to every integer n=1,2,… we assign a set. The ordered class of sets $A_1, A_2, \ldots$ a sequence $\{A_n\}$ of sets.

Increasing sequence of sets: Let $\{A_n\}$ be a sequence of sets. If, every set of $\{A_n\}$ is a subset of its succeeding set i.e. $A_n \subseteq A_{n+1}, \ \forall n$, then $\{A_n\}$ is known as monotonically increasing set.

In this case we denote $\{A_n\} \uparrow$. **In this case** $\bigcup_{k=1}^{n} A_k = A_n; \bigcup_{k=1}^{\infty} A_k = A$ **is called the limit of** $\{A_n\}$. **Symbolically** $A_n \uparrow A$.

Evidently A contains all the A_n 's and is the smallest of such sets.

Decreasing sequence of sets: Let $\{A_n\}$ be a sequence of sets. If, every set of $\{A_n\}$ is a subset of its preceding set i.e. $A_n \supseteq A_{n+1}, \ \forall n$, then $\{A_n\}$ is known as monotonically decreasing sequence of sets. In this case we denote $\{A_n\}\downarrow$. **In this case** $\bigcap_{k=1}^{n} A_k = A_n; \bigcap_{k=1}^{\infty} A_k = A$ **is called the limit of** $\{A_n\}$ **. Symbolically** $A_n \downarrow A.$.

Evidently A is contained in all the A_n 's and is the largest of such sets.

Thus, if the sequence $\{A_n\}$ is non-decreasing or non-increasing, the limit exists and we have

$$\lim_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} A_n \qquad \text{if } \{A_n\}\uparrow$$

and

$$\lim_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} A_n \qquad \text{if } \{A_n\}\downarrow$$

Limit Infimum: Let $\{A_n\}$ be a sequence of sets. The set of all points that belongs to A_n for all but a finite number of values of *n* is known as the limit infimum (or) limit inferior of the sequence $\{A_n\}$ and is denoted by

$$\lim_{n\to\infty} \inf A_n \text{ or } \underline{\lim} A_n = \{\omega : \omega \in A_n \text{ for all } A_n \text{ except perhapes } A_1, A_2, \ldots, A_n, \text{ but a finite number of n}\}$$

Note: We have $\underline{\lim}_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k \subseteq \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k = \overline{\lim}_{n\to\infty} A_n$

Limit Supermom: Let $\{A_n\}$ be a sequence of sets. The set of all those elements which belong to A_n for infinitely many values of ***n*** is known as the limit supermom (or) limit superior of the sequence, usually denoted by

$$\lim_{n\to\infty} \sup A_n \text{ or } \overline{\lim} A_n = \{x : x \in A_n \text{ for all,infintely many n}\}$$

$$\lim_{n\to\infty} \sup A_n \text{ or } \overline{\lim} A_n = \{x : x \in A_n \text{ for all,infintely many n}\}$$

Limit of arbitrary sequence of sets: Let $\{A_n\}$ an arbitrary sequence of sets. For any sequence $\{A_n\}$ define

$$B_n = \inf_{k\ge n} A_k = \bigcap_{k=n}^{\infty} A_k = \{\omega : \omega \in A_n \text{ for all } A_n \text{ except perhapes } A_1, A_2, \ldots, A_{n-1}\}$$

$$C_n = \sup_{k\ge n} A_k = \bigcup_{k=n}^{\infty} A_k = \{\omega : \omega \text{ belong to at least one of } A_n, A_{n+1}, \ldots\},$$

In particular, $B_1 = \bigcap_{k=1}^{\infty} A_k$, $B_2 = \bigcap_{k=2}^{\infty} A_k$

Therefore, $B_1 = A_1 \cap \left(\bigcap_{k=2}^{\infty} A_k \right) = A_1 \cap B_2 \subseteq B_2$

$B_1 \subset B_2$

Similarly, we can show that $B_2 \subseteq B_3$, $B_3 \subseteq B_4$, ..., $B_{n-1} \subseteq B_n$.

Thus, $\{B_n\}$ is an increasing sequence of sets and hence

$$\lim_{n\to\infty} \{B_n\} \text{ is } \bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k = \underline{\lim}\ A_n$$

Thus B_n is a monotonically increasing sequence with limit $B = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k = \lim\inf A_k = \underline{\lim}\ A_n$.

B is the set of all points which belong to almost all A_n (a;; but any finite number of sets).

Similarly, C_n a monotonically decreasing sequence with limit

$$C = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k = \lim\sup A_k = \overline{\lim}\ A_n \ .$$

C is the set of all those points which belongs to infinitely many A_n .

Since every point which belongs to almost all A_n belongs to infinitely many A_n,

$\underline{\lim}\ A_n \subseteq \overline{\lim}\, A_n$.

If $\underline{\lim}\ A_n = \overline{\lim}\, A_n = A$ **(say), the limit of** $\{A_n\}$ **is said to exist and** A **is called the limit of** $\{A_n\}$ **.**

1.8 CONCLUSION:

A **field** (algebra) is closed under finite unions, finite intersections, and complements, but not necessarily countable operations. A **σ-field** (σ-algebra) extends a field by also being closed under countable unions and intersections. The **minimal field** containing a collection of sets is the smallest field including those sets. The **minimal σ-field** containing a collection of sets is the smallest σ-field including them, which is critical in probability theory (e.g., Borel σ-algebra).

1.9 SELF ASSESSMENT QUESTIONS:

1. Define an **event** in probability theory. Give an example of an event in a real-life scenario.

2. Define a **sequence of sets** and give an example.

3. Given a class $\{A_i, i = 1, 2, ..., n\}$ of n sets prove that there exists a class $\{B_i, i = 1, 2, ..., n\}$ of n disjoint sets such that their unions are equal.

4. Define minimal field. Explain the procedure of obtaining a minimal filed over a class σ-filed. Extend this procedure to arrive at the minimal σ-filed.

5. Explain limit Supremum and limit infimum of sequence of sets.

1.10 SUGGESTED READING BOOKS:

1) Modern probability theory by B. R. Bhat, Wiley EasternLimited.
2) An introduction to probability theory and mathematical statistics by V. K. Rohatgi, John Wiley.
3) An Outline of statistics theory-1, by A.M.GOON, M.K. Gupta and B. Das gupta, the World Press Private Limited, Calcutta.
4) The Theory of Probability by B.V. Gnedenko, MIR Publishers, Moscow.
5) Discrete distributions -N.L. Johnson and S. Kotz, John wiley & Sons.
6) ContinuousUnivariatedistributions,vol.1&2N.L.JohnsonandS.Kotz,JohnWiley&Sons.
7) Mathematical Statistics-Parimal Mukopadhyay, New Central Book Agency (P) Ltd., Calcutta.

Dr. Syed Jilani

LESSON -2

AXIOMATIC DEFINITION OF PROBABILITY AND CONTINUITY THEOREM OF PROBABILITY

OBJECTIVES :

After studying this unit, you should be able to:

- To understanding the axiomatic definition of probability.
- To know the concept of Structure and axiomatic definition of probability.
- To acquire knowledge about significance of axiomatic definition of probability.
- To understand the purpose and objectives of pivotal provisions of the axiomatic definition of probability.

STRUCTURE:

2.1 INTRODUCTION:

Probability is a mathematical framework used to quantify uncertainty and measure the likelihood of events occurring. The **axiomatic definition of probability** was introduced by **Andrey Kolmogorov** in 1933 and is based on a set of fundamental principles known as **Kolmogorov's Axioms**. These axioms provide a rigorous foundation for probability theory and are widely used in mathematics, statistics, and various applications.

The **Continuity Theorem of Probability**, sometimes called the **Continuity Property**, extends the idea of additivity from finite collections of events to infinite collections. It helps us understand the behavior of probabilities for sequences of events that either increase or decrease.

2.2 SAMPLE SPACE AND CLASS OF EVENTS:

Sample Space:

Def.1: A random experiment is an experiment in which

(a) All outcomes of experiment are known in advance.

(b) Any performance of experiment results in an outcome that is not is not known in advance.

(c) The experiment can be repeated under identical conditions.

In probability theory we study this uncertainty of a random experiment. It is convenient to associate with each such experiment a set Ω,the set of all possible outcomes of experiment. To engage in any meaningful discussion about the experiment, we associate with ca σ-field S of subsets of Ω .we recall that a σ-field is a nonempty class of subset of Ω that closed under formation of countable unions and complements and contains the null set $\varnothing$.

Def 2: the sample space of statistical experiment is a pair (Ω , S),where

(a) Ω is the set all possible of the experiment.

(b) S is a σ-field of subsets of Ω .

The element of Ω are call sample points. Any set $A \in$ S is known as an event. Clearly, A is a collection of sample points. We say that an event A happens if the outcome of the experiment corresponds to a point in A. Each one point set is known as a simple or elementary event. if the set Ω . Contains only a finite number of points, we say that (Ω , S) is a finite sample space. If Ω .contains at most a countable number of points, we call (Ω , S) a discrete sample space. If however, Ω . Contains uncountably many points, we say that (Ω , S)is an uncountable sample space. In particular, if $\Omega = \Re_k$ or some rectangle in $\Re_k$,We call it a continuous sample space,

REMARK 1: The choices of S is an important one, and some remarks are in order. If Ω contains at most a countable number of points, we can always take S to be the class of all subsets of Ω.this certainly a C Each one point set is a member of S and is the fundamental objet of interest, Every subset of Ω is an event, If Ω has uncountably many points ,the class of all subsets of Ω is still a σ-field, but it is much to large a class of sets to be interest .one of the most important examples uncountable sample space is the case in which $\Omega = \Re$ or Ω is an interval in $\Re$ .in this case we would like all one-point subsets of Ω and all interval to be events. We use our knowledge of analysis to specify S . We will not go into detail here expert to recall that the class of all semi closed intervals (a,b] generates a class B_1 that is on $\Re$ .this class contains all one point sets and all intervals. we take S=B_1 .Since we will be dealing mostly with the one-dimensional case. We write B instead of B_1. there are many subsets of $\Re$ that are not in B_1 ,but we do not demonstrate this fact here.

Example 1: Let us toss a coin. The set Ω is the set of symbols H and T, where H denotes head and T represents tail. Also, S is the class of all subsets of Ω,namely, $\{\{H\},\{T\},\{H,T\},\varnothing\}$ .If the coin is tossed two times, then

$$\Omega=\{\{H,H\},\{H,T\},\{T,H\}\{\mathrm{T,T}\}\}$$

And

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

Where the first element of a pair denotes the outcome of the first toss, and the second element, the outcome of the second toss. The event at least one head consist of sample points $(H,H),(H,T),(T,H)$. The event at most one head is the collection of sample points $(H,T),(T,H),(T,T)$.

Example 2: A die is rolled n-times. The sample space is the pair (Ω,S), where Ω is the set of all n-tuples $(x_1,x_2,...,x_n), x_i\in\{1,2,3,4,5,6\}, i=1,2,..,n,$ and S is the class of all subjects of Ω. Ω contains 6^n elementary events. The event A that 1 shows at least once is he set

$$A=\{(x_1,x_2,...,x_n): \textit{at least once of } x_i^{'}s \textit{ is } 1\}$$
$$=\Omega-\{(x_1,x_2,...,x_n): \textit{none once of } x_i^{'}s \textit{ is } 1\}$$
$$=\Omega-\{(x_1,x_2,...,x_n): x_i\in\{2,3,4,5,6\}, i=1,2,..,n\}.$$

Example 3: A coin is tossed until the first head appears. Then $\Omega=\{H,(T,H),(T,T,H),(T,T,T,H),...\}$, and S is the class of all subsets of Ω. An equivalent way of writing Ω would be to look at the number of toss required for the first head. Clearly, this number can take values 1,2,..., so that Ω is the set of all positive integers. Thus S is the class of all subset of positive integers.

Class of Events: An event is defined as collection of outcomes of an experiment E, in which one is interested. Since Ω is defined as the collection of all outcomes of E, if A is an event, A^c will also be an event. It denotes non-occurrence of A. If A is the class f all events associated with the experiment E, it is closed under complementation.
If A and B are events, then $A\cup B$ denotes the event that at least one of the events A or B occurs; $A\cap B$ denotes the event that both the events A and B occur simultaneously. Thus A is closed under finite unions and intersections and is a field. If $A_1, A_2,..., A_n$ are events $\bigcup_{i=1}^{n} A_i$ will denote the events that at least one event A will occur; A_i will denote the event that all the events A_i occur simultaneously. Since the union and intersection can be extended to the case when n is infinite, A will naturally be a $\sigma-\textit{field}$ if E has an infinite number of outcomes.

Thus, with each experiment E, we can associate Ω, the space of all outcomes and A, the $\sigma-\textit{field}$ of events. For different experiments we shall have different Ω and different A.

This become a handicap when we have to consider outcomes of more than one experiment simultaneously. This handicap is avoided by considering Ω to be the collection of outcomes of all experiments "under consideration." However A, the $\sigma-\textit{field}$ of events, will be associated with only a particular experiment in which one is interested, For example, Ω may consists of a dice-throwing and a coin-tossing experiment. The events of interest may be based on the outcomes of coin-tossing experiment only. In this case an event is a subset of Ω and not conversely.

With every experiment having a finite number of possible outcomes, the class of all events will be a field. If the field is degenerate, the only events are ϕ and Ω . These are called respectively as impossible and sure events in line with our proposal to associate chance with the events in the later chapters. If the number of possible outcomes is infinite, then the class of events is a $\sigma - field$.

Definition: the space Ω of all outcomes of experiments together with the specification of the $\sigma - field \;\; A$ of events of sometimes called a measurable space and is denoted by the pair (Ω, A) . Any set of A is called a measurable set.

Let A be a class of events and $B \in A$. Then for $A \in A$, $B \bigcap A$ is an event. The class of all such events

$$\left\{ B \bigcap A : A \in A \right\} = B \bigcap A, say$$

Is a $\sigma - field$ of subsets of B . It is the restriction of A to subsets of B and hence possesses the same properties as those of A .

Here are some examples across different contexts:

1. Business & Economics

- Financial Crises (e.g., 2008 Global Financial Crisis, Asian Financial Crisis)
- Market Trends **(**e.g., Bull Market, Bear Market)
- Corporate Mergers & Acquisitions
- Product Launches

2. Technology & Cybersecurity

- Cyberattacks **(**e.g., Phishing, Ransomware, DDoS Attacks)
- Software Updates & Releases
- System Failures & Outages
- AI Breakthroughs

3. Science & Nature

- Natural Disasters **(e.g.,** Earthquakes, Hurricanes, Wildfires)
- Climate Change Events (e.g., Rising Sea Levels, Heatwaves)
- Space Exploration Missions
- Epidemics & Pandemics

4. Social & Cultural

- Protests & Movements (e.g., Civil Rights Movement, Climate Strikes)
- Major Sports Events (e.g., Olympics, FIFA World Cup)
- Award Ceremonies (e.g., Oscars, Grammys)
- Political Elections

5. Military & Conflict

- Wars & Battles (e.g., World War I, Gulf War)
- Peace Treaties & Agreements
- Terrorist Attacks
- Military Operations

2.3 PROBABILITY MEASURE OR AXIOMATIC DEFINITION OF PROBABILITY:

Let (Ω, S) be the sample space associated with a statistical experiment. Now, we define a probability set function and study some of its properties.

Definition: Let (Ω, S) be the sample pace. A set function P defined on S is called a probability measure (or simply, probability) if it satisfies the following conditions:

i. $P(A) \geq 0, \forall A \in \mathrm{S}$

ii. $P(\Omega) = 1$

iii. Let $\{A_j\}, A_j \in \mathrm{S}, j = 1, 2, \ldots,$ be a disjoint sequence of sets; that is, $A_j \bigcap A_k = \phi$ for $j \neq k,$ where ϕ is the null set. Then

$$P\left(\sum_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} P(A_j),$$

where we have used the notation $\sum_{j=1}^{\infty} A_j$ to denote union of disjoint sets A_j.

We call $P(A)$ the probability of the event A. If there is no confusion, we will write PA instead $P(A)$. Property iii is called countable additivity. That $P\phi = 0$ and P is also finitely additive follows from it.

Remark 1: If Ω is discrete and contains at most $n(<\infty)$ points, each single point set $\{\omega_j\}, j = 1, 2, \ldots, n,$ is an elementary event, and it is sufficient to assign probability each $\{\omega_j\}$. Then if $A \in S$, S is the class all students of Ω, $PA = \sum_{\omega \in A} P\{\omega\}$. One such assignment is the equally likely assignment or the assignment of uniform probabilities. According to this assignment, $P\{\omega_j\} = 1/n, j = 1, 2, \ldots, n.$ Thus $PA = m/n$ if A contains m elementary events, $1 \leq m \leq n.$

Remark 2: If Ω is discrete and contains a countable number of points, one cannot make an equally likely assignment of probabilities. It suffices to make the assignment for each elementary event. If $A \in S$, where S is the class of all subsets of Ω, define $PA = \sum_{\omega \in A} P\{\omega\}.$

Remark 3: If Ω contains uncountable many point, each one-point set is an elementary event, and again one cannot make an equally likely assignment of probabilities. Indeed, One cannot assign positive probability to each elementary event without violating the axiom $P(\Omega) = 1.$ In this case one assign probabilities to compound events consisting of intervals. For example, if

$\Omega = [0,1]$ and S is the Borel σ-field of all subjects of Ω, the assignment $P[I] =$ length of I, where I is the subinterval of Ω, defines a probability.

Remark 4: The triple (Ω, S, P) is called a probability space.

Remark 5: Let $A \in S$. We say that the odds for A are a to b if $PA = a/(a+b)$, and then odds against A are b to a.

Note: If 's' contains only a finite no. of sets. Then it will be sufficient. If $P(.)$ satisfies 1,2,3 only

- In general if S is only a field of events. Then $P(.)$ defined on S need be only finite additive.
- If S is a σ-field then $P(.)$ has to be σ-additive

Now the triplet (Ω, S, P) where Ω is the space of outcomes, S is the σ-field of events associated with an experiment and P is a probability function defined on S is called as a probability space.

Discrete probability space:

In a sample space Ω if the class of events S is generated by countable partition of subset of Ω then (Ω, S, P) is called the "*discrete probability space*".

- If Ω is countable A_i could be singletons $\{\omega_i\}(i = 1,2,...)$ Now the set $\{p_1, p_2, ...\}$ where $p_i = P(\omega_i)$ is called the probability distribution on Ω then S co-insides with the power set.
- If Ω is finite and if $p_i = P(\omega_i)$ is the same for all $i = 1,2,...,N$ then this corresponds to the equally likely case with (1/N, 1/N,… 1/N) as the probability distribution.

2.4 DEFINITION OF MEASURE:

Let Ω be the space of all possible outcomes and S is the σ-field defined on Ω. Now a real valued function μ [is said to be a measure] defined on S is called as a measure if for any $A_1, A_2, ... \in S$, μ statistics the properties.

I. $\mu(A_i) \geq 0$ (non-negative)

II. If $A = \sum_{i=1}^{n} A_i$ then $\mu(A) = \sum_{i=1}^{n} \mu(A_i)$ (finite additivity)

Note:

- If $\mu(\Omega)$ is finite then μ is said to be a **finite measure**. **Thus probability is a special case of finite measure**. We may also say that probability is a normed or scaled measure.
- All the general properties of measures will be shared by probability measure. It possess some original properties. Some additional property namely $P(\Omega) = 1$.
- If $\mu(\Omega) = \infty$ but $\Omega = \sum_{i=1}^{\infty} A_I$ such that $\mu(A_i) < \infty$, then μ is called a σ-finite measure.

  Measures such as length defined on (R,B) is a σ-finite measure know the triplet (Ω, S, μ) is called the measure space.

2.5 CONTINUITY THEOREM OF PROBABILITY FOR MONOTONICALLY NON-DECREASING SEQUENCE OF EVENTS:

Statement: If $\{A_n\}$ is a monotonically non-decreasing sequence of events in a probability space (Ω, S, P) whose limit is say A, then show that

$$P(\mathrm{A}) = P\left(\lim_{n\to\infty} \mathrm{A}_n\right) = \lim_{n\to\infty} P(\mathrm{A}_n) \tag{1}$$

Proof:

Since $\{A_n\}$ is a monotonically non-decreasing sequence of events, we have

$$\mathrm{A}_n \subseteq \mathrm{A}_{n+1}, \quad \forall n \tag{2}$$

and by definition we have

$$\lim_{n\to\infty} \mathrm{A}_n = \mathrm{A} = \bigcup_{n=1}^{\infty} \mathrm{A}_n \tag{3}$$

Let us define

$$B_n = \mathrm{A}_n - \mathrm{A}_{n-1}, \quad \forall n \tag{4}$$

so that B_n's are disjoint and from Eq.(4), we have

$$B_1 = \mathrm{A}_1$$

$$\mathrm{B}_1 U \mathrm{B}_2 = \mathrm{A}_1 U (\mathrm{A}_2 - \mathrm{A}_1) = \mathrm{A}_1 U \mathrm{A}_2 = \mathrm{A}_2 \qquad \text{(From Eq.(2))}$$

$$\mathrm{B}_1 U \mathrm{B}_2 U \mathrm{B}_3 = \mathrm{A}_1 U \mathrm{A}_2 U (\mathrm{A}_3 - \mathrm{A}_2) = \mathrm{A}_1 U \mathrm{A}_2 U \mathrm{A}_3 = \mathrm{A}_3 \qquad \text{(From Eq.(2))}$$

….

$$\bigcup_{i=1}^{n} B_i = \bigcup_{i=1}^{n} \mathrm{A}_i = \mathrm{A}_n \tag{5}$$

From the above we may also write

$$\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n = \sum_{n=1}^{\infty} B_n \qquad \text{(since } B_n\text{'s are disjoint)} \tag{6}$$

Now, from Eqs.(3) and (6), we may write

$$\mathrm{A} = \lim_{n\to\infty} \mathrm{A}_n = \sum_{n=1}^{\infty} B_n$$

Now applying probability function on both sides and by using σ-additivity axiom of probability, we have

$$P(\mathrm{A}) = P\left(\lim_{n\to\infty} \mathrm{A}_n\right) = P\left(\sum_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} P(B_n) = \lim_{n\to\infty} \sum_{i=1}^{n} P(B_i) \tag{7}$$

From Eq. (5), we have

$$\mathrm{A}_n = \bigcup_{i=1}^{n} B_i = \sum_{i=1}^{n} B_i \qquad \text{(since } B_i\text{'s are disjoint)}$$

Applying probability function on both sides and by using finite- additivity axiom of probability, we have

$$P(\mathrm{A}_n) = P\left(\sum_{i=1}^{n} B_i\right) = \sum_{i=1}^{n} P(B_i) \tag{8}$$

From Eqs. (7) and (8), we get

$$P(\mathrm{A}) = P\left(\lim_{n\to\infty} \mathrm{A}_n\right) = \lim_{n\to\infty} \sum_{i=1}^{n} P(B_i) = \lim_{n\to\infty} P(\mathrm{A}_n) \tag{9}$$

Hence the result (1).

i.e. The probability set function and limit of sequence are interchangeable.

2.5.1 CONTINUITY THEOREM OF PROBABILITY FOR MONOTONICALLY NON-INCREASING SEQUENCE OF EVENTS:

Statement: If $\{A_n\}$ is a monotonically non-increasing sequence of events in a probability space (Ω, S, P) whose limit is say A, then show that

$$P(\mathrm{A}) = P\left(\lim_{n\to\infty} \mathrm{A}_n\right) = \lim_{n\to\infty} P(\mathrm{A}_n) \tag{1}$$

Proof:

Since $\{A_n\}$ is a monotonically non-increasing sequence of events, we have

$$\mathrm{A}_n \supseteq \mathrm{A}_{n+1}, \quad \forall n \tag{2}$$

and by definition we have

$$\lim_{n\to\infty} \mathrm{A}_n = \bigcap_{n=1}^{\infty} \mathrm{A}_n = \mathrm{A} \quad \text{(say)} \tag{3}$$

Since $\{A_n\}$ is a monotonically non-increasing sequence of events, we have $\{\overline{A}_n\}$ is a monotonically non-decreasing sequence of events and from the above Continuity theorem of Probability of monotonically non-decreasing sequence of events, it follows immediately

$$P\left(\lim_{n\to\infty} \overline{\mathrm{A}}_n\right) = \lim_{n\to\infty} P(\overline{\mathrm{A}}_n) \tag{4}$$

Since $\{\overline{A}_n\}\uparrow$, we have

$$\lim_{n\to\infty} \overline{\mathrm{A}}_n = \bigcup_{n=1}^{\infty} \overline{\mathrm{A}}_n = \overline{\left(\bigcap_{n=1}^{\infty} \mathrm{A}_n\right)} = \overline{\mathrm{A}} \qquad \text{(from Eq. (3))} \tag{5}$$

Upon using Eq.(5) in Eq. (4), we have

$$P(\overline{\mathrm{A}}) = \lim_{n\to\infty} P(\overline{\mathrm{A}}_n) \tag{6}$$

$$\Rightarrow 1 - P(\mathrm{A}) = \lim_{n\to\infty}\left[1 - P(\mathrm{A}_n)\right] = 1 - \lim_{n\to\infty} P(\mathrm{A}_n)$$

$$\Rightarrow P(\mathrm{A}) = \lim_{n\to\infty} P(\mathrm{A}_n) \tag{7}$$

Now, from Eqs.(3) and (7), we have

$$\Rightarrow P(\mathrm{A}) = P\left(\lim_{n\to\infty} \mathrm{A}_n\right) = \lim_{n\to\infty} P(\mathrm{A}_n)$$

Hence the result (1).

i.e. The probability set function and limit of sequence are interchangeable

2.5.2 CONTINUITY THEOREM OF PROBABILITY FOR AN ARBITRARY SEQUENCE OF EVENTS, WHOSE LIMIT EXISTS:

Statement: If $\{A_n\}$ is an arbitrary sequence of events in a probability space (Ω, S, P) has a limit is say A, then show that

$$P(\mathrm{A}) = P\left(\lim_{n\to\infty} \mathrm{A}_n\right) = \lim_{n\to\infty} P(\mathrm{A}_n) \tag{1}$$

Proof:

We have $$\bigcap_{i=n}^{\infty} \mathrm{A}_i \subseteq \mathrm{A}_n \subseteq \bigcup_{i=n}^{\infty} \mathrm{A}_i \Rightarrow P\left(\bigcap_{i=n}^{\infty} \mathrm{A}_i\right) \le P(\mathrm{A}_n) \le P\left(\bigcup_{i=n}^{\infty} \mathrm{A}_i\right) \tag{2}$$

If we write $B_n = \bigcap_{i=n}^{\infty} \mathrm{A}_i$ and $\mathrm{C}_n = \bigcup_{i=n}^{\infty} \mathrm{A}_i$, then $\{B_n\}\uparrow$ & $\{\mathrm{C}_n\}\downarrow$ (3)

and from Eq. (2), we can write $P(B_n) \le P(\mathrm{A}_n) \le P(C_n)$

Taking limits we get $$\lim_{n\to\infty} P(B_n) \le \lim_{n\to\infty} P(\mathrm{A}_n) \le \lim_{n\to\infty} P(C_n) \tag{4}$$

But, from the continuity theorems of probability for monotonic non-decreasing and non-increasing sequences of events, we have

$$\lim_{n\to\infty} P(B_n) = P\left(\lim_{n\to\infty} B_n\right) \quad \text{and} \quad \lim_{n\to\infty} P(C_n) = P\left(\lim_{n\to\infty} C_n\right) \tag{5}$$

Upon using Eq.(5) in Eq.(4), we get $$P\left(\lim_{n\to\infty} B_n\right) \le \lim_{n\to\infty} P(\mathrm{A}_n) \le P\left(\lim_{n\to\infty} C_n\right) \tag{6}$$

And by definition, we have

$$\lim_{n\to\infty} B_n = \bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty}\bigcap_{i=n}^{\infty} \mathrm{A}_i = \underline{\lim}\ \mathrm{A}_n \qquad \text{(by def. of limit infimum)}$$

$$\text{and } \lim_{n\to\infty} \mathrm{C}_n = \bigcap_{n=1}^{\infty} \mathrm{C}_n = \bigcap_{n=1}^{\infty}\bigcup_{i=n}^{\infty} \mathrm{A}_i = \overline{\lim}\ \mathrm{A}_n \qquad \text{(by def. of limit suprimum)} \tag{7}$$

Now, upon using Eq. (7) in Eq.(6), we get $$P(\underline{\lim}\ \mathrm{A}_n) \le \lim_{n\to\infty} P(\mathrm{A}_n) \le P\left(\overline{\lim}\ \mathrm{A}_n\right) \tag{8}$$

But, from the hypothesis, $\lim_{n\to\infty} \mathrm{A}_n$ exists and therefore by definition, we have

$$\mathrm{A} = \lim_{n\to\infty} \mathrm{A}_n = \underline{\lim}\ \mathrm{A}_n = \overline{\lim}\ \mathrm{A}_n$$

Using this in Eq. (8), we get

$$P(\mathrm{A}) \le \lim_{n\to\infty} P(\mathrm{A}_n) \le P(\mathrm{A}) \Rightarrow P(\mathrm{A}) = P\left(\lim_{n\to\infty} \mathrm{A}_n\right) = \lim_{n\to\infty} P(\mathrm{A}_n) \tag{9}$$

Hence the result (1).

Remark1: Thus, σ- additivity condition is sometimes referred to as the continuity condition of the probability function. The notions of continuity from below, continuity from above and continuity in general, are similar to the notions of continuity from the left, from the right and in the absolute sense for point functions.

Remark2: Even though the above theorems are proved for probability measures, results contained therein continue to hold for finite measures with obvious modification in the proof. If the measure is σ-finite, it will be continuous from above but we cannot establish continuity from below.

2.6 CONDITIONAL PROBABILITY:

Introduction:

So far we have computed probabilities of events on the assumption that no information was available about the experiment other than the sample space Ω. Sometimes however it is that an event H which is a subset of Ω i.e., $(\mathrm{H} \subseteq \Omega)$ has happened. Now using this information making a probability statement concerning the outcomes of another event A is called conditional probability.

Definitions:

Let $(\Omega, \mathrm{S}, \mathrm{P})$ be a probability space and let $\mathrm{H} \in \mathrm{S}$ with $\mathrm{P}(\mathrm{H}) > 0$. Now for an arbitrary event $\mathrm{A} \in \mathrm{S}$ we shall write $\mathrm{P}(\mathrm{A}/\mathrm{H}) = \dfrac{\mathrm{P}(\mathrm{A} \cap \mathrm{H})}{\mathrm{P}(\mathrm{H})}$ and call the so defined quantity as the conditional probability of A/H i.e., $\mathrm{P}(\mathrm{A}/\mathrm{H})$. Conditional probability remains undefined when $\mathrm{P}(\mathrm{H}) = 0$.

Result:

Let $(\Omega, \mathrm{S}, \mathrm{P})$ be a probability space and let $\mathrm{H} \in \mathrm{S}$ with $\mathrm{P}(\mathrm{H}) > 0$ then $(\Omega, \mathrm{S}, P_H)$ where $P_H(\mathrm{A}) = \mathrm{P}(\mathrm{A}/\mathrm{H}) \forall \mathrm{A} \in \mathrm{S}$ is probability space.

Proof:

(a) Clearly $P_H(\mathrm{A}) = \mathrm{P}(\mathrm{A}/\mathrm{H}) \geq 0 \forall \mathrm{A} \in \mathrm{S}$

(b) Also $P_H(\Omega) = \mathrm{P}(\Omega/\mathrm{H}) = \dfrac{\mathrm{P}(\Omega \cap \mathrm{H})}{\mathrm{P}(\mathrm{H})} = \dfrac{\mathrm{P}(\mathrm{H})}{\mathrm{P}(\mathrm{H})} = 1$

(c) If $A_1, A_2, \ldots$ is a disjoint sequence of sets in S then

$$P_H\left(\sum_{\mathrm{i}=1}^{\infty} A_i\right) = \mathrm{P}\left(\sum_{\mathrm{i}=1}^{\infty} A_i/\mathrm{H}\right)$$

$$= \frac{\mathrm{P}\left(\sum_{\mathrm{i}=1}^{\infty} A_i/\mathrm{H}\right)}{\mathrm{P}(\mathrm{H})}$$

But we know that from $\sigma-$ additivity axiom of probability we have

$$\mathrm{P}\left(\sum_{\mathrm{i}=1}^{\infty} A_i \cap \mathrm{H}\right) = \sum_{\mathrm{i}=1}^{\infty} \mathrm{P}(A_i \cap \mathrm{H})$$

$$\therefore P_H\left(\sum_{\mathrm{i}=1}^{\infty} A_i\right) = \frac{\sum_{\mathrm{i}=1}^{\infty} \mathrm{P}(A_i \cap \mathrm{H})}{\mathrm{P}(\mathrm{H})}$$

$$= \sum_{\mathrm{i}=1}^{\infty} \mathrm{P}(A_i/\mathrm{H})$$

$$= P_H(A_i)$$

Hence the third axiom

$\therefore (\Omega, \zeta, P_H)$ is a probability space.

Remark:

What we have done is we consider a new sample space consisting of a basic set H and the $\sigma-$ field $S_H = \zeta \cap H$ of subsets $A \cap H \, \forall A \in S(H)$.

On the space we have defined a set function P_H by multiplying the probability of each event by $\frac{1}{P(H)}$.

Indeed (H, S_H, P_H) is a probability space.

Results:

1) Let A and B be two events with $P(A) > 0, P(B) > 0$ then from the definition of conditional probability we have

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

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

2) The above result can be extended for three results A,B,C where $P(A) > 0, P(B) > 0$ and $P(c) > 0$ we have

$$\begin{aligned} P(A \cap B \cap C) &= P(A/B \cap C) \cdot P(B \cap C) \\ &= P(A/B \cap C)P(B/C)P(C) \\ &= P(A)P(B/A)P(C/A \cap B) \end{aligned}$$

3) For any arbitrary n events $A_1, A_2, \ldots., A_n$,

$$P\left(\bigcap_{i=1}^{n} A_i\right) = P(A_1)P(A_2/A_1)P(A_3/A_1 \cap A_2), \ldots., P\left(A_n \Big/ \bigcap_{i=1}^{n-1} A_i\right)$$

2.7 CONCLUSION:

The study of probability theory begins with a precise understanding of the **sample space** and the **class of events**. The sample space Ω\OmegaΩ represents the set of all possible outcomes of an experiment, while the class of events (often taken as a σ\sigmaσ-algebra) consists of the subsets of Ω\OmegaΩ on which probability measures can be defined. This careful structuring is essential for a rigorous treatment of probability.

Building on this foundation, the **axiomatic definition of probability** (via Kolmogorov's axioms) provides a formal framework in which probability is defined as a function PPP that assigns a number between 0 and 1 to each event in the sigma -algebra. The three axioms—non-negativity, normalization, and countable additivity—ensure that probabilities are consistent and behave in an intuitive manner. These axioms serve as the basis for all further developments in probability theory.

In parallel, the **definition of measure** generalizes the idea of probability to more abstract settings. A measure is a function defined on a σ -algebra that satisfies properties analogous to the axioms of probability, particularly countable additivity. This abstraction is critical for integrating probability with analysis and for handling spaces where classical notions of "length" or "volume" may not apply directly.

2.8 SELF ASSESSMENT QUESTIONS:

1. Discuss the properties derived from the axiomatic definition of probability.
2. Explain why probability values are always between 0 and 1 using axiomatic principles.
3. Define a measure on a measurable space and explain the concept of a sigma-algebra.
4. How does a probability measure differ from a general measure? Provide an example.
5. Define conditional probability.
6. Explain the Continuity Theorem of Probability for monotonically non-decreasing sequence of Events.
7. Explain the Monotonically non-increasing sequence of Events and arbitrary sequence of Events.
8. Explain the Continuity Theorem of Probability for an arbitrary sequence of Events, whose limit exists.

2.9 SUGGESTED READING BOOKS:

1. Modern probability theory by B. R. Bhat, Wiley EasternLimited.
2. An introduction to probability theory and mathematical statistics by V. K. Rohatgi, John Wiley.
3. AnOutlineofstatisticstheory-1, by A.M.GOON, M.K. Gupta and B. Das gupta, the World Press Private Limited, Calcutta.
4. The Theory of Probability by B.V. Gnedenko, MIR Publishers, Moscow.
5. Discrete distributions -N.L. Johnson and S. Kotz, John wiley & Sons.
6. ContinuousUnivariatedistributions,vol.1&2N.L.JohnsonandS.Kotz,JohnWiley&Sons.
7. Mathematical Statistics-Parimal Mukopadhyay, New Central Book Agency (P) Ltd., Calcutta.

Dr. Syed Jilani

LESSON -3

BOREL CANTELLI LEMMA AND GEOMETRIC PROBABILITY

OBJECTIVES :

After studying this unit, you should be able to:

- To understanding the Statistical independence of events and Geometric probability.
- To know the concept of Structure and Statistical independence of events and Geometric probability.
- To acquire knowledge about significance of Statistical independence of events and Geometric probability.
- To understand the purpose and objectives of pivotal provisions of the Statistical independence of events and Geometric probability.

STRUCTURE:

3.1 INTRODUCTION:

Probability theory provides powerful tools to analyze the occurrence of events over repeated trials. Two important topics in this field are the **Borel-Cantelli Lemma**, which helps determine whether an infinite sequence of events will occur infinitely often, and **Geometric Probability**, which applies probability theory to spatial problems.

3.2 STATISTICAL INDEPENDENCE OF EVENTS

We shall now introduce the notion of independence of events in suppose P(A)>0. If B be another event such that $P(B\ /\ A) = P(B)$ (1)

Then the probability of the occurrence of B would remain unaffected by the knowledge that A has occurred. We may then say that B is (statistically or probabilistically) independent of A since

$$P\left({}^{B}/_{A}\right) = \frac{P(A \cap B)}{P(A)}$$

Equation (2.20) is equivalent to $P(A \cap B) = P(A)P(B) \ldots (2)$ indeed we may say that B is independent of A, by definition, if $P(A \cap B) = P(A)P(B) \ldots$ (3)

Again , since Eq(1) is symmetrical in A and B. one should speak about the mutual independence of A and B rather than about A being independent of B or B being independent of A. Also relation Eq (1) is regarded as the definition of the mutual independence of A and B irrespective of whether $P(A)$ or$P(A^C)$ or $P(B)$ or$P(B^C)$ is zero or not although in that case the conditional probabilities will not be all defined.

Theorem:1

Let $A \sqsubseteq \mathbb{C}, B \sqsubseteq \mathbb{C}$ then if A and B are independent so are

(1)A and B^C

(2) A^C and B

(3) A^C and B^C

Proof: let A and B be independent then

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

Since $P(A \cap B) + P(A \cap B^C) = P(A)$

This implies $P(A \cap B^C) = P(A) - P(A \cap B)$

$$= P(A) \quad P(A)P(B)$$

$$= P(A)[1 - P(B)]$$

$$- P(A)P(B^C)$$

Hence A and B^C are independent

(2) This can be proved in the same way as (1) has been proved

(3) Again let A and B be independent then

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

Hence $P(A^C \cap B^C) = 1 - P(A \cup B)$

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

$$= \{1 - P(A)\}\{1 - P(B)\}$$

$$= P(A^C)P(B^C)$$

Thus A^C and B^C are independent if A and B are independent

Example 1:

Let the four letters a, b, c and d be written down in random order. Then let A and B denote respectively, the events "a proceeds b" and "c proceeds d" the total number of such equations is

$$(r_2) + (r_3) + \cdots + (r_r) - (1+1)^r - (r_0) - (r_1)$$

$$= 2^r - r - 1$$

Incidentally , for the mutual independence of r events $(r > 2)$, it is not enough that the events be pair-wise independent. This may be illustrated by taking the simple case at three events, say $A_1, A_2 and A_3$.

Example2:

Let us suppose that for an experiment the sample space consists of four points only : $\omega_1, \omega_2, \omega_3 and \omega_4$. Let $P(\{\omega\}) = P_i = {}^1/_{4,} for\ i = 1,2,3,4$ as would be the case in two throws of a perfect coin.

Consider three events $A_1, A_2, and\ A_3$ defined as follows:

$$A_1 = \{\omega_1, \omega_2\}, A_2 = \{\omega_1, \omega_3\}\ and\ A_3 = \{\omega_1, \omega_4\}$$

Then $P(A_1) = P(A_2) = P(A_3) = {}^1/_2$

And $P(A_1 \cap A_2) = {}^1/_4 = P(A_1)P(A_2)$,

$$P(A_1 \cap A_3) = {}^1/_4 = P(A_1)P(A_3)$$

$$P(A_2 \cap A_3) = {}^1/_4 = P(A_2)P(A_3)$$

So that the three events are pair-wise independent. But for mutual independence of the events it is also necessary that $P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3)$. How ever, in the present case,

$$P(A_1 \cap A_2 \cap A_3) = {}^1/_4$$

While $P(A_1)P(A_2)P(A_3) = {}^1/_8$

This counter-example is due to Bernstein

*If P(A) or P(B) is zero, then (2.21) is necessarily satisfied, for $P(A \cap B) \leq P(A)$ and $P(A \cap B) \leq P(B)$, and so A and B are then taken to be independent.

Here the total number of elementary events in 4! and these are equally likely

Again , the number of elementary events favorable to A and the number favorable to A^C must be the same, by symmetry , and so each is 12. Similarly , the number of elementary events favorable to B is 12. On the other hand , the elementary events favourable to both A and B are

The following six:

(a,b,c,d),(a,c,b,d),(a,c,d,b)(c,d,a,b),(c,a,d,b)

Hence P(A)=P(B) = $^{12}/_{24} = ^{1}/_{2}$

While $P(A \cap B) = ^{6}/_{24} = ^{1}/_{4}$

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

And so the events are statistically independent

In the general case of r events, $A_1, A_2, \ldots . A_r$ we say that these are mutually independent if each is independent ofall and any of the others. The definition of mutual independence of these events is given in terms of the following equations.

$$P(A_i \cap A_j) = P(A_i)P(A_j)\, for\ 1 \leq i \leq j \ \leq r;$$

$$P(A_i \cap A_j A_k) = P(A_i)P(A_j)P(A_k)\, for\ 1 \leq i \leq j \ \leq k \leq r;$$

. . . . .

. . . . .

. . . . .

$$P(A_1 \cap A_2 \cap \ldots \cap A_r) = P(A_1)P(A_2) \ldots P(A_r) \ldots .$$

3.3 BOREL CANTELLI LEMMA:

Statement: If $\{A_n\}$ is a countable collection of events and A_n is the limit supremum of $\{A_n\}$ then show that

(a) $\sum_{n=1}^{\infty} P(A_n) < \infty$ (i.e. convergent) $\Rightarrow P(A) = 0$ (1)

(b) A_n's are independent and $\sum_{n=1}^{\infty} P(A_n) = \infty$ (i.e. divergent) $\Rightarrow P(A) = 1$ (2)

Proof:

(a) Since, A is the limit supremum of the sequence $\{A_n\}$, by definition, we have

$$\mathrm{A} = \overline{\lim}\ \mathrm{A}_n = \bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} \mathrm{A}_i$$

We know that , $\mathrm{A} = \bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} A_i \subseteq \bigcup_{i=n}^{\infty} A_i$

$$\Rightarrow P(A) = P\left(\bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} A_i\right) \le P\left(\bigcup_{i=n}^{\infty} A_i\right) \tag{3}$$

But from addition theorem of probability, we have

$$P\left(\bigcup_{i=n}^{\infty} A_i\right) \le \sum_{i=n}^{\infty} P(A_i) \tag{4}$$

From Eqs. (3) and (4), it immediately follows

$$P(A) \le \sum_{i=n}^{\infty} P(A_i)$$

Taking limit on both sides, we get

$$\lim_{n\to\infty} P(A) \le \lim_{n\to\infty} \sum_{i=n}^{\infty} P(A_i)$$

$$i.e. \quad P(A) \le \lim_{n\to\infty} R_n \tag{5}$$

$$\text{where } \; R_n = \sum_{i=n}^{\infty} P(A_i) \tag{6}$$

Let us define $S_n = \sum_{i=1}^{n-1} P(A_i)$ so that $\sum_{i=1}^{\infty} P(A_i) = S_n + R_n$ (7)

Since $\sum_{i=1}^{\infty} P(A_i)$ is convergent, we have $\sum_{i=1}^{\infty} P(A_i) < \infty$. Now, let us write

$$\sum_{i=1}^{\infty} P(A_i) = l \quad (say) \tag{8}$$

so that from Eqs. (7) & (8), we may write

$$l = S_n + R_n$$

Taking limit on both sides we get,

$$l = \lim_{n\to\infty} S_n + \lim_{n\to\infty} R_n \tag{9}$$

But, from Eq.(7), we have

$$\lim_{n\to\infty} S_n = \lim_{n\to\infty} \sum_{i=1}^{n-1} P(A_i) = \sum_{i=1}^{\infty} P(A_i) = l \tag{10}$$

Thus, from Eqs.(9) & (10),

$$\lim_{n\to\infty} R_n = 0 \quad \Rightarrow P(A) \le 0 \qquad \text{(from Eq.(5))}$$

$$\Rightarrow P(A) = 0 \quad i.e. \quad P\left(\lim_{n\to\infty} A_n\right) = 0$$

Hence the proof of **(a).**

(b) we have given

$$\mathrm{A} = \overline{\lim} A_n = \bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} A_i \Rightarrow \overline{\mathrm{A}} = \bigcup_{n=1}^{\infty} \bigcap_{i=n}^{\infty} \overline{\mathrm{A}}_i$$

$$\Rightarrow P(\overline{\mathrm{A}}) = P\left(\bigcup_{n=1}^{\infty} \bigcap_{i=n}^{\infty} \overline{\mathrm{A}}_i\right) = P\left(\lim_{n\to\infty} \bigcap_{i=n}^{\infty} \overline{\mathrm{A}}_i\right)$$

$$\Rightarrow P(\overline{\mathrm{A}}) = \lim_{n\to\infty} P\left(\bigcap_{i=n}^{\infty} \overline{\mathrm{A}}_i\right) \tag{11}$$

(By continuity theorem for monotonically incresing sequence of sets)

For N>n, we see that

$$\bigcap_{i=n}^{\infty} \overline{\mathrm{A}}_i \subseteq \bigcap_{i=n}^{N} \overline{\mathrm{A}}_i \Rightarrow P\left(\bigcap_{i=n}^{\infty} \overline{\mathrm{A}}_i\right) \le P\left(\bigcap_{i=n}^{N} \overline{\mathrm{A}}_i\right) \tag{12}$$

Taking $\lim\limits_{N\to\infty}$ on both sides we get,

$$P\left(\bigcap_{i=n}^{\infty} \overline{\mathrm{A}}_i\right) \le \lim_{N\to\infty} P\left(\bigcap_{i=n}^{N} \overline{\mathrm{A}}_i\right)$$

Since, A_i's are independent, it is obviously $\overline{\mathrm{A}}_i$'s are also independent. Therefore, by virtue of Multiplication theorem, we have

$$P\left(\bigcap_{i=n}^{\infty} \overline{\mathrm{A}}_i\right) \le \lim_{N\to\infty} P\left(\bigcap_{i=n}^{N} \overline{\mathrm{A}}_i\right) = \lim_{N\to\infty} \prod_{i=n}^{N} P(\overline{\mathrm{A}}_i) = \lim_{N\to\infty} \prod_{i=n}^{N} \left[1 - P(\mathrm{A}_i)\right] \tag{13}$$

For any $0 \le \alpha_i \le 1,$ we have

$$e^{-\alpha_i} = 1 - \alpha_i + \frac{\alpha_i^2}{2!} - \frac{\alpha_i^3}{3!} + \ldots \ge 1 - \alpha_i$$

$$\left\{ \text{since } \frac{\alpha_i^2}{2!} - \frac{\alpha_i^3}{3!} = \frac{\alpha_i^2}{2!}\left(1 - \frac{\alpha_i}{3}\right) \ge 0,\ \frac{\alpha_i^4}{4!} - \frac{\alpha_i^5}{5!} = \frac{\alpha_i^4}{4!}\left(1 - \frac{\alpha_i}{5}\right) \ge 0, \text{ and so on} \right\}$$

$$\Rightarrow 1 - \alpha_i \le e^{-\alpha_i} \Rightarrow \prod_{i=n}^{N} (1 - \alpha_i) \le \prod_{i=n}^{N} e^{-\alpha_i} = e^{-\sum_{i=n}^{N} \alpha_i} \tag{14}$$

Since, $0 \le P(\mathrm{A}_i) \le 1,$ we may take $\alpha_i = P(\mathrm{A}_i)$ in Eq. (14), and then from Eq. (13), it follows

$$P\left(\bigcap_{i=n}^{\infty} \overline{\mathrm{A}}_i\right) \le \lim_{N\to\infty} \prod_{i=n}^{N}\left[1 - P(\mathrm{A}_i)\right] \le \lim_{N\to\infty} e^{-\sum_{i=n}^{N} P(\mathrm{A}_i)} = e^{-\sum_{i=n}^{\infty} P(\mathrm{A}_i)} \tag{15}$$

Since, the series $\sum_{i=1}^{\infty} P(A_i)$ is divergent, we have $\sum_{i=n}^{\infty} P(\mathrm{A}_i) = \infty$. Therefore, from Eq.(15), we have

$$P\left(\bigcap_{i=n}^{\infty} \overline{\mathrm{A}}_i\right) \le e^{-\sum_{i=n}^{\infty} P(\mathrm{A}_i)} = e^{-\infty} = 0 \tag{16}$$

Now, from Eqs. (11) and (16), we may write

$$P(\overline{\mathrm{A}}) = \lim_{n\to\infty} P\left(\bigcap_{i=n}^{\infty} \overline{\mathrm{A}}_i\right) \le \lim_{n\to\infty} e^{-\infty} = 0 \tag{17}$$

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

Hence, the proof of **Part(b).**

3.4 PROBABILITY ON FINITE SAMPLE SPACES:

In this we restrict attention to sample spaces that have at most a finite number of points. Let $\Omega = \{\omega_1, \omega_2, ...\omega_n\}$ and S be the σ-field of all subsets of Ω. For any $A \in S$,

$$P(A) = \sum_{\omega \mathrm{j} \in A} P\{\omega_j\}$$

Example 1: A coin is tossed twice. The sample space consists of four points. Under the uniform assignment, each of four elementary events is assigned probability $\frac{1}{4}$.

Example 2: Three dices are rolled. The sample space consists of 6^3 points. Each one –point set is assigned probability $\frac{1}{6^3}$.

In games of chances we usually deal with finite sample spaces where uniform probability is assigned to all simple events. The same is the case in sampling schemes. In such instances the computation of probability of event A reduces to a combinatorial counting problem. We therefore consider some rules of counting.

Rule 1: Given a collection of n_1 elements $a_{11}, a_{12}, \ldots a_{1n1}, n_2$ elements $a_{21}, a_{22}, \ldots a_{2n2}$, and so on, up to n_k elements $a_{k1}, a_{k2}, \ldots a_{kn_k}$, it is possible to form $n_1, n_2 \ldots n_k$ ordered k-tuples $\left(a_{1\,j_1}, a_{2\,j_2}, \ldots a_{k\,j_k}\right)$ containing one element of each kind, $1 \le j_i \le n_i$, $i = 1, 2, \ldots k$.

Rule 2 is concerned with ordered samples. Consider a set of n elements $a_1, a_2 \ldots a_n$. Any ordered arrangement $\left(a_{i1}, a_{i2} \ldots, a_{ir}\right)$ of r of these n symbols is called *ordered sample of size* r. If elements are selected one by one, there are two possibilities:

1. Sampling with replacement: In this case repetitions are permitted, and we can draw sample of an arbitrary size. Clearly, there n^r sample size r.

2. Samping without replacement: In these case an element once chosen is not replaced, so that there can be no repetitions. Clearly the sample size cannot be exceed n, the size of population. There are $n(n-1) \ldots (n-r+1) = {}_nP_r$, say, possible samples of size r. Clearly, ${}_nP_r = 0$ for integers $r > n$. If $r = n$, then ${}_nP_r = n!$

Remark 1: We frequently use the term random sample in this book to describe the equal assignment of probability to all possible samples in sampling from a finite population. Thus , when we speak of a random sample of size r from a population of a n elements , it means that in sampling with replacement, each of n^r sample has the same probability $\frac{1}{n^r}$ or that in sampling without replacement, each of ${}_nP_r$ samples is assigned probability $\frac{1}{{}_nP_r}$.

Rule 3: There are $\binom{n}{r} = \frac{n!}{r!(n-r)!}$ different subpopulations of size $r \le n$ from a population of n elements, where

Rule 4: Consider a population of n elements. The number of ways in which the population can be partition into k subpopulation of size $r_1, r_2 \ldots r_k$, respectively $r_1 + r_2 + \ldots + r_k = n$, $0 \le r_i \le n$, is given by

$$\binom{n}{r_1, r_2 \ldots, r_k} = \frac{n!}{r_1!\, r_2! \ldots r_k!}$$

The numbers define in above are known as *multinomial coefficients.*

Example 3: In a game of bridge the probability that a hand of 13 cards contains 2 spades, 7 hearts, 3 diamonds, and 1 club is

$$\frac{\binom{13}{2}\binom{13}{7}\binom{13}{3}\binom{13}{1}}{\binom{52}{13}}$$

Example 4: An urn contains 5 red, 3 green, 2 blue, and 4 white balls. A sample of size 8 is selected at random without replacement. The probability that the sample contains 2 red, 2 green, 1 blue, and 3 white balls is

$$\frac{\binom{5}{2}\binom{3}{2}\binom{2}{1}\binom{4}{3}}{\binom{14}{8}}$$

3.5 GEOMETRIC PROBABILITY:

Geometrical approach to the calculation of probabilities is employed when the sample space Ω includes an uncountable set of elementary event ω and none of them is more likely to occur than the other. Suppose as in the sample space Ω is a domain in a plane and the elementary events ω are points within Ω. If an event A is represented by the event A within Ω, so that all ω belonging to the region A is are favorable to the event A, than probability of A is

$$p(A) = \frac{area \text{ of subdomain } A}{area \text{ of domain } \Omega} \tag{1}$$

In general, Probability of event $A = \dfrac{S_A}{S_\Omega}$ (2)

Where S_A and S_Ω are measures of the specified part of the region representing the event A and the sample space Ω reprehensively. The measure may length, area, volume etc. According as region is in one, two, and three,...dimensions.

Formula (1) is a generalization of classical to an uncountable set of elementary events. The symmetry of the experimental conditions with respect to the elementary outcomes ω is usually formulated by the assumption of randomness.

The theory of geometric probabilities is often criticized for arbitrariness is determining the probability of events. Many authors point out that for an infinite number of outcomes the probability cannot determined objectively i.e. independently of the mode of computation.

Example 1: A straight line of unit length say, (0, 1) is divided into three intervals by choosing two points at random. What is the probability that the three line segments form a triangle ?

Let x, y be the abscissas of any two points chosen on (0,1). A set of necessary and sufficient conditions for the three segments to form a triangle is that the sum of the other two. Hence we should have either

$$0 < x < \frac{1}{2} < y < 1 \quad \text{and} \quad y - x < \frac{1}{2}$$

(Or) $$0 < y < \frac{1}{2} < x < 1 \quad \text{and} \quad x - y < \frac{1}{2}$$

This is represented by the shaded area in the figure. Hence the probability is $\frac{1}{4}$.

Example 2: Buffon's needle problem. A plane is ruled with paralled straight line at distance L from each other. A needle of length $l < L$ is thrown at random on the plane. Find the probability that it will hit one of the lines.

We characterize the outcome of the experiment by two numbers: the abscissa x of the centre of the needle with respect to the left and by the angle θ the needle makes with the direction of the lines. Since the needle is thrown at random all values of x and θ are equiprobable. Without any loss of generality we consider only the possibilities of the needle hitting the nearest line on the left when $0 \le x \le \frac{L}{2}$ and $0 \le \theta \le \frac{\pi}{2}$ (The probability is same for hitting the nearest line on the right). The sample space Ω is thus a rectangle of area $S_\Omega = \frac{L\pi}{4}$. The needle will be hit the line if $x < \frac{l}{2}\sin\theta$. We are thus interested in the event $A = \left\{ x < \frac{l}{2}\sin\theta \right\}$. Area S_A of $A = \int_0^{\frac{\pi}{2}} \frac{l}{2}\sin\theta \, d\theta$.

Hence $$p(A) = \frac{S_A}{S_\Omega} = \frac{2l}{\pi L} = \frac{2 \text{ length of the needle}}{\text{circumference of a circle of radius } \frac{L}{2}}.$$

[This is an interesting problem, perhaps, because it suggests a relation between a pure chance experiment and a famous number π. If we take a graph paper ruled by parallel line 1 inch apart and a length 1 inch and keep track of the fraction of the times the needle crosses a line, when thrown randomly on the graph paper, π may be estimated as about 2/(proportion of crosses).]

Example 3: If a chord is selected at random on fixed circle what is the probability that its length exceeds the radius of the circle?

Depending on the manner how the term 'random' is explained, we have three different answers. Let the radius of the circle be r.

(a) Assume that the distance of the chord from the centre of the circle is a random value within 0 to r. A regular hexagon *abcdef* of side r can be inscribed in the circle. Any chord lying within this polygon will have length gather than r. Hence the required probability $= \frac{h}{r}$, h being the distance of ab from O. Now $h = r\sqrt{\frac{3}{2}}$. Hence probability $= \frac{\sqrt{3}}{2} \approx 0.866$.

(b) Assume that the midpoint of the chord is evenly distributed over the interior of the circle. The chord is longer than the radius when the midpoint of the chord is within h

of the centre. Thus all points in the circle of the radius h with o as centre can within h of the centre serve as the mid-point of the chord. Hence probability $= \dfrac{\pi h^2}{\pi r^2} = \dfrac{3}{4}$.

(c) Assume that the chord is determined by two points chosen so that their positions are independently evenly distributed over the circumference of the circle. Assuming that the first point falls at a , the second point must fall on the *arc fab* for the chord to the shorter then radius. The length of *fab* = $\frac{1}{3}$ of the circumference.

Hence probability = $\frac{2}{3}$.

3.6 CONCLUSION:

The study of probability theory, particularly in the context of statistical independence, the Borel-Cantelli Lemma, probability on finite sample spaces, and geometric probability, provides a deep understanding of how events occur and interact in different settings.

(i) **Statistical Independence of Events**: Understanding independence is crucial for modeling real-world phenomena where events occur without influencing each other. It is a fundamental concept in probability, helping in simplifying complex probability computations.

(ii) **Borel-Cantelli Lemma**: This lemma plays a significant role in understanding whether a sequence of events will happen infinitely often. It is widely used in probability theory and applications like number theory, statistical physics, and stochastic processes.

(iii) **Probability on Finite Sample Spaces**: Finite sample spaces provide a basis for understanding probability computations in controlled environments. The difference between **sampling with replacement** and **sampling without replacement** affects the probability distribution of outcomes and is important in statistics and machine learning.

(iv) **Geometric Probability**: The application of probability to geometric settings allows for analyzing spatial randomness, such as random points, lines, and shapes in a given space. Problems like Buffon's needle or random chord distributions provide insight into real-world stochastic processes.

3.7 SELF ASSESSMENT QUESTIONS:

1. Define statistical independence of two events and provide an example.
2. What is geometric probability? Provide an example where it is applied.
3. Prove that if two events are independent, their complements are also independent.
4. Explain the second Borel-Cantelli Lemma and describe a situation where it applies.
5. Consider a scenario where a fair die is rolled five times. Compute the probability of getting at least one six:
   a) If rolls are independent (with replacement).
   b) If rolls are dependent (without replacement in a different context, e.g., drawing numbered balls).
6. Explain the application of geometric probability in physics and engineering. Provide examples such as random scattering of particles or fiber alignment in composite materials.

3.8 SUGGESTED READINGS:

1) Modern probability theory by B. R. Bhat, Wiley EasternLimited.
2) An introduction to probability theory and mathematical statistics by V. K. Rohatgi, John Wiley.
3) An Outline of statistics theory-1, by A.M.GOON, M.K. Gupta and B. Das gupta, the World Press Private Limited, Calcutta.
4) The Theory of Probability by B.V. Gnedenko, MIR Publishers, Moscow.
5) Discrete distributions -N.L. Johnson and S. Kotz, John wiley & Sons.
6) ContinuousUnivariatedistributions,vol.1&2N.L.JohnsonandS.Kotz,JohnWiley&Sons.
7) Mathematical Statistics-Parimal Mukopadhyay, New Central Book Agency (P) Ltd., Calcutta.

Dr. Syed Jilani

LESSON -4

RANDOM VARIABLES

OBJECTIVES:

After studying this unit, you should be able to:

- To understanding the Random Variable
- To know the concept of Structure and Random Variable
- To acquire knowledge about significance of Random Variable
- To understand the purpose and objectives of pivotal provisions of the Random Variable

STRUCTURE:

4.1. INTRODUCTION:

In **LESSON**-I we dealt essentially with random experiment that can be described by finite sample spaces. We studied the assignment and computation of probabilities of events. In practice, one observes a function defined on the space of outcomes. Thus, if a coin is tossed n times, one is not interested in knowing which of the 2^n n-tuples in the sample space has occurred. Rather, one would like to know the number of heads in n tosses. In games of chance, one is interested in the net gain or loss of a certain player. Actually, in lesson -I we were concerned with such functions without defining the term random variable. Here we study the notion of a random variable and examine some of its properties.

In **LESSON** -I we studied properties of a set function P defined on a sample space (Ω, S). Since P is a set function, it is not very easy to handle; we cannot perform arithmetic or algebraic operations on sets. Moreover, in practice one frequently observes some function of elementary events. When a coin is tossed repeatedly, which replication resulted in heads is not of much interest. Rather, one is interested in the number of heads, and consequently, the number of tails, which appear in, say, n tossings of the coin. It is therefore desirable to introduce a point function on the sample space. We can then use our knowledge of calculus or real analysis to study properties of P .

4.2 RANDOM VARIABLE:

Definition: Let (Ω, S) be a sample space. A finite, single-valued function that maps Ω into $\mathcal{R}$ is called a <u>random variable</u> (RV) if the inverse images under X of all Borel sets in $\mathcal{R}$ are events, that is, if

$$X^{-1}(B) = \{\omega : X(\omega) \in B\} \in S \quad \text{for all } B \in \mathcal{B}. \tag{1}$$

To verify whether a real- valued function on (Ω, S) is an RV, it is not necessary to check that (1) holds for all Borel sets $B \in \mathcal{B}$. It suffices to verify (1) for any class □ of subsets of $\mathcal{R}$ that generates $\mathcal{B}$. By taking □ to be the class of semi-closed intervals $(-\infty, x]$, $x \in \mathcal{R}$. We get the following result.

Result: X is a random variable $\Leftrightarrow \forall x \in \mathcal{R}, \{\omega : X(\omega) \le x\} = \{X \le x\} \in S$ (2)

Remarks:

1. Note that the notion of probability does not enter into the definition of an R.V.
2. If X is an RV, the sets $\{X = x\}$, $\{a < X \le b\}, \{X < x\}, \{a \le X < b\}$, $\{a < X < b\}$, $\{a \le X \le b\}$ are all events. Moreover, we could have used any of these intervals to define an RV. For example, we could have used the following equivalent definition: X is an RV if and only if

   $$\{\omega : X(\omega) < x\} \in S \quad \forall x \in \Re$$

   We have $\{X < x\} = \bigcup_{n=1}^{\infty}\left(X \le x - \frac{1}{n}\right)$ and $\{X \le x\} = \bigcap_{n=1}^{\infty}\left(X < x + \frac{1}{n}\right)$
3. In practice, (1) or (2) is a technical condition in the definition of an RV which the reader may ignore and think of RVs simply as real-valued functions defined on Ω. It should be emphasized, though, that there do exist subsets of $\mathcal{R}$ that do not belong to $\mathcal{B}$, and hence there exist real-valued functions defined on Ω that are not RVs, but the reader will not encounter them in practical applications.
4. If X is a random variable, the sets $\{X = x\}, \{a < X \le b\}$ $\{Xx\}, \{a \le X < b\}, \{a \le X \le b\}$ are all events.
5. In practice, a random variable is simply a real valued function defined on a Sample Space Ω .

Example: Let $\Omega = \{HH, TT, HT, TH\}$ and S be the class of all subsets of Ω. Define X by $X(\omega) =$ number of H's in ω.

Then $X(HH) = 2, X(HT) = X(TH) = 1,$ and $X(TT) = 0$. Then

$$X^{-1}(-\infty, x] = \begin{cases} \phi, x < 0 \\ \{TT\}, 0 \le x < 1 \\ \{TT, TH, HH\}, 1 \le x < 2 \\ \Omega, x \ge 2 \end{cases}$$

Then X is a random variable.

For example $X^{-1}(-\infty, 1.5) = \{TT, TH, HT\}$.

Some more Examples:

1. Let X be the number of tails in the tosses of a coin. What is Ω? What are the values that X assigns to points of Ω? What are the events

   $\{X \le 2.75\}, \{0.5 \le X \le 1.72\}$

2. A die is tossed two times. Let X be the sum of face values on the tosses and Y be the absolute value of the differences in the face values. What is Ω? What values do X and Y assign to points of Ω? Check to see whether Xand Y are random variables.
3. A die is rolled 5 times. Let X be the sum of face values. write the sets {x=4},{x=6}, {x=30}, and {x≥29}.

4.2.1 Probability distribution of a random variable:

For understanding the concept of a random variable, in practice, we define it only on a sample space (Ω, S). However, random variables are of interest only when they are defined on a probability space. Let (Ω, S, P) be a probability space and let X be a random variable defined on it.

Theorem: The R.V. X defined on the probability space (Ω, S, P) induces a probability space $(\mathscr{R}, \mathscr{B}, Q)$ by means of the correspondence

$Q(B) = P\{X^{-1}(B)\} = P\{\omega \in \Omega / x(\omega) \in B\}$ for all $B \in \mathscr{B}$ we write Q= PX^{-1} and call Q or PX^{-1} as the probability distribution of X.

Proof: Clearly, $Q(B) = P\{\omega \in \Omega / X(\omega) \in B\} \ge 0$ for all $B \in \mathscr{B}$.

Also $\mathbf{Q}(\mathscr{R}) = P\{\omega \in \Omega / X(\omega) \in \mathscr{R}\} = P(\Omega) = 1$

Let $B_i \in \mathscr{B}$, i=1,2,… with $B_i \bigcap B_j = \phi, i \ne j$ .i.e. $B's$ are disjoints.

Since, the inverse image of a disjoint union of Borel sets is the disjoint union of these inverse images, we have.

$$\mathbf{Q}\left(\sum_{i=1}^{\infty} B_i\right) = P\left\{X^{-1}\left(\sum_{i=1}^{\infty} B_i\right)\right\} = P\left\{\sum_{i=1}^{\infty} X^{-1}(B_i)\right\} = \sum_{i=1}^{\infty} PX^{-1}(B_i) = \sum_{i=1}^{\infty} \mathbf{Q}(B_i)$$

i.e. additivity of probability is proved. Thus, it follows $(\mathscr{R}, \mathscr{B}, Q)$ is a probability space. Hence the proof.

4.3 DISTRIBUTION FUNCTION:

Distribution function: A real valued function F defined on $(-\infty,\infty)$ that is non-decreasing, right continuous and satisfies

$F(-\infty)=0$ and $F(+\infty)=1$

is called a distribution function (DF).

Alternative definition: Let 'X' be a random variable on (Ω, S, P). Then the function

$$F(x)=P(X\le x)=P\{\omega : X(\omega)\le x\}, -\infty<x<\infty$$

is called the distribution function of a random variable X.

Properties of Distribution Function:-

We now proceed to derive a number of properties common to all distribution functions.

Property 1:-

If F is the distribution function of the random variable X and if $a<b$, then

$$P(a<X\le b)=F(b)-F(a)$$

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

$$P(a<X\le b)+P(X\le a)=P(X\le b)$$

$$P(a<X\le b)=P(X\le b)-P(X\le a)=F(b)-F(a)$$

Corollary 1:-

$$P(a<X\le b)=P\{(X=a)\cup(a<X\le b)\}=P(X=a)+P(a<X\le b)$$

(Using additive theorem of p)

$$=P(X=a)+[F(b)-F(a)]$$

Similarly, we get

$$P(a<X<b)=P(a<X\le b)-P(X=b)$$

$$=F(b)-F(a)-P(X=b)$$

$$P(a\le X<b)=P(a<X<b)+P(X=a)$$

$$=F(b)-F(a)-P(X=b)+P(X=a)$$

Remark: when $P(X=a)=0$ and $P(X=b)=0$, all four events $a\le X\le b, a<X<b, a\le X<b$ and $a<X\le b$, have the same Probability $F(b)-F(a)$.

Property 2:-

If F is the distribution function of one –dimensional random variable X, then,

1. $0 \le F(x) \le 1$
2. $F(x) \le F(y)$ if x<y

In other words, all distribution functions are monotonically non-decreasing and lie between 0 and 1.

Proof: - Using the axioms of certainty and non- negativity for the probability function P, part (1) follows triviality from the definition of $F(x)$.

For part (2), we have for $x < y$,

$$F(y) - F(x) = P(x < X \le y) \ge 0 \Rightarrow F(y) \ge F(x) \Rightarrow F(x) \le F(y) \text{ when } x < y$$

Property 3: If F is distribution function of one- dimensional random variable x, then

$$F(-\infty) = \lim_{x \to -\infty} F(x) = 0 \quad \text{and} \quad F(\infty) = \lim_{x \to \infty} F(x) = 1$$

Proof: - let us express the whole sample space S as a countable union of disjoint events as follows:

$$S = \left[\bigcup_{n=1}^{\infty} (-n < X \le -n+1)\right] \cup \left[\bigcup_{n=0}^{\infty} (n < X \le n+1)\right]$$

$$\Rightarrow P(S) = \sum_{n=1}^{\infty} P(-n < X \le -n+1) + \sum_{n=0}^{\infty} P(n < X \le n+1) \quad \text{(P is additive)}$$

$$\Rightarrow 1 = \lim_{a \to \infty} \sum_{n=1}^{a} \left[F(-n+1) - F(-n)\right]$$

$$\Rightarrow 1 = \lim_{a \to \infty} \sum_{n=1}^{a} \left[F(-n+1) - F(-n)\right] + \lim_{b \to \infty} \sum_{n=0}^{b} \left[F(n+1) - F(n)\right]$$

$$= \lim_{a \to \infty} \left[F(0) - F(-a)\right] + \lim_{b \to \infty} \left[F(b+1) - F(0)\right]$$

$$= \left[F(0) - F(-\infty)\right] + \left[F(\infty) - F(0)\right]$$

$$1 = F(\infty) - F(-\infty)$$

Since $-\infty < \infty, F(-\infty) \le F(\infty)$. Also

$$F(-\infty) \ge 0 \text{ And } F(\infty) \le 1$$

$$0 \le F(-\infty) \le F(\infty) \le 1 \qquad (2)$$

1 and 2 give $F(-\infty) = 0$ and $F(\infty) = 1$

REMARKS:-

1. Discontinuous of F(x) are at most countable
2. $F(a) - F(a-0) = \lim_{h \to 0} P(a - h \le X \le a), h > 0$

$$F(a) - F(a-0) = P(X = a)$$

And $\quad F(a+0) - F(a) = \lim_{h \to 0} P(a \le X \le a+h) = 0, h > 0$

$$\Rightarrow F(a+0) = F(a)$$

Theorem: The function F defined by $F(x) = \mathrm{P}\{\omega \in \Omega / \mathrm{X}(\omega) \le x\}, \forall x \in \mathrm{R}$ is indeed a DF.

Proof: Let $x_1 \le x_2$, then we have, $(-\infty, x_1] \subset (-\infty, x_2)$

$$\Rightarrow P\{(-\infty, x_1]\} \le P\{(-\infty, x_2)\}$$

$$\Rightarrow \mathrm{P}\{\omega \in \Omega / \mathrm{X}(\omega) \le x_1\} \le \mathrm{P}\{\omega \in \Omega / \mathrm{X}(\omega) \le x_2\}$$

$$\Rightarrow \mathrm{P}\{X \le x_1\} \le \mathrm{P}\{X \le x_2\}$$

$$\Rightarrow F(x_1) \le F(x_2) \tag{1}$$

Thus F is non-decreasing function.

In order to show F is right continuous, we consider sequences of numbers.

$\{x_n\} \downarrow x$ i.e; $x_1 > x_2 > \ldots > x_n > \ldots > x$

Let $A_k = \{\omega \in \Omega / X(\omega) \in (x_1, x_k]\}$ (2)

Thus $A_k \in \mathbf{S}$, sample space and $\{A_k\}$ is non-increasing. Since, none of the intervals $(x_1, x_k]$ contains x,

We have,

$$\lim_{k \to \infty} A_k = \bigcap_{k=1}^{\infty} A_k = \phi \tag{3}$$

Taking probability on both sides we have

$$P\left(\lim_{k \to \infty} A_k\right) = P(\phi) = 0$$

$\Rightarrow \lim_{k \to \infty} P(A_k) = 0$ (From the continuity property of P) (4)

But,

$$P(A_k) = P\{(\omega \in \Omega / X(\omega) \in (x, x_k])\}$$

$$= P\{(\omega \in \Omega / X(\omega) \in (-\infty, x_k])\} - P\{(\omega \in \Omega / X(\omega) \in (-\infty, x])\}$$

$$= P(X \le x_k) - P(X \le x) = F(x_k) - F(x) \tag{5}$$

Using (4) and (5), we get, $\lim_{k \to \infty} \{F(x_k) - F(x)\} = 0$

$$\Rightarrow \lim_{k \to \infty} F(x_k) = F(x)$$

$\Rightarrow F$ is right continuous.

Finally, Let $\{x_n\}$ be a sequence of numbers decreasing to $-\infty$.

Then

$$x_{n+1} \le x_n \Rightarrow \{X \le x_{n+1}\} \subseteq \{X \le x_n\}, \forall n \Rightarrow \{X \le x_n\} \supseteq \{X \le x_{n+1}\}, \forall n$$

$$\Rightarrow \lim_{n\to\infty}\{X \le x_n\} = \bigcap_{n=1}^{\infty}\{X \le x_n\} = \phi$$

Taking probability on both sides, we get $P\left(\lim_{n\to\infty}\{X \le x_n\}\right) = P(\phi) = 0$

But, $P\left(\lim_{n\to\infty}\{X \le x_n\}\right) = \lim_{n\to\infty} P\{X \le x_n\} = \lim_{n\to\infty} F(x_n) = F(-\infty)$

$\therefore F(-\infty) = P(\phi) = 0$

Similarly, we can prove, $F(+\infty) = \lim_{n\to\infty} P(X \le x_n) = 1$

Thus, F is non-decreasing, right continuous, $F(-\infty) = 0$ and $F(+\infty) = 1$

$\therefore$ F is indeed a DF.

Theorem: The distribution function F of RV X is non-decreasing, continuous on the right with $F(-\infty) = 0$ and $F(+\infty) = 1$. Conversely, every function F, with the above properties is the distribution function of a RV on some probability space.

Proof: Same as the above

Distribution function

Let X be a real random variable on the probability space (Ω, S, P) . For $x \in R$, define,

i. $P[X \le x] = F(x)$,
ii. $P(a, b\ ] = P(a < x \le b)\ = F(b) - F(a), (b > a)$

This $F(x)$ is called the distribution function (d.f.) or cumulative distribution of X.

4.3.1 Discrete random variable:

Definition 1: A random variable X defined on (Ω, S, P) is said to be of discrete type, or simply discrete if there exists a countable set $P\{X \in E\} = 1$. The points of E that have positive mass are called 'jump points' or points of increase of the distribution function of X, and their probabilities are called jumps of the distribution function.

Definition 2: The collection of numbers $\{p_i\}$ satisfying $P\{\omega \in \Omega / X(\omega) = x_i\} = p_i \ge 0, \forall i$ and $\sum_{i=1}^{\infty} p_i = 1$. is called the probability mass function (p.m.f) of random variable X.

The distribution function F of X is given by

$$F(x) = P[X \le x] = \sum_{x_i \le x} p_i$$

A Result: Let $\{p_i\}$ be a collection of non-negative real numbers such that $\sum_{k=1}^{\infty} p_k = 1$. Then $\{p_k\}$ is the p.m.f. of some RV X.

Examples:

1. $\sum_{k=0}^{\infty} \frac{e^{-\lambda}\lambda^k}{\underline{|k}} = 1, \quad \left(e^{-\lambda}e^{\lambda}\right)$

   Solution $\left\{p_k = \frac{e^{-\lambda}\lambda^k}{\underline{|k}}\right\}$ is the p.m.f. of Poisson distribution.

2. $\sum_{x=0}^{\infty} pq^x = 1, \quad \left(p(1-q)^{-1}\right)$

   $$\sum_{x=0}^{\infty} pq^x = p.\left(1+q+q^2+...\right)$$

   $$= p(1-q)^{-1} = \frac{p}{1-q} = \frac{p}{p} = 1$$

   Solution $\{pq^x\}, x = 0,1,2,...$ is the p.m.f. of geometric distribution.

3. $\sum_{x=0}^{n} \binom{n}{x} p^x q^{n-x} = (p+q)^n = 1$

   Solution $\left\{\binom{n}{x} p^x q^{n-x}\right\}$ is the p.m.f. of Binomial distribution.

4.3.2 Continuous random variable:

Let X be a random variable defined on (Ω, S, P) with distribution function F. Then X is said to be continuous, if F is absolutely continuous, that is if there exists a non-negative function $f(x)$ such that

$$F(x) = \int_{-\infty}^{\infty} f(t)dt, \forall x \in R.$$

The function f is called the probability density function (p.d.f.).

Result 1: Every non-negative real function f that is integrable on R and satisfying

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

is the p.d.f. of some continuous random variable X.

Result 2: $F'(x) = \frac{d}{dx}F(x) = f(x).$

Define the distribution function of **two dimensional R.V**'s show that the necessary and sufficient conditions for it to be the distribution function of a two dimensional random variable.

Let X be a r.v defined on (Ω, S, P) defined a function $\mathrm{F}(.)$ on R namely

$$\mathrm{F}(x) = \mathrm{Q}(-\infty, x]$$

$$= \mathrm{P}\{\omega \in \Omega / \mathrm{X}(\omega) \subseteq x\} \forall x \in \mathrm{R}$$

Now the function F is called the D.F of r.v "X"

i. F is non decreasing and right continuous with respect to both arguments.

ii. $F(-\infty, y) = F(x, -\infty) = 0$ and $F(-\infty, +\infty) = 1$

iii. For every $(x_1, y_1)(x_2, y_2)$ with $x_1 < x_2$ and $y_1 < y_2$ the inequality (2) $F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1) - F(x_1, y_2) \geq 0$ holds

i. Proof: For $x < x'$

$\{X \leq x'\} = \{X \leq x\} + \{x < X \leq x'\}$

Taking probability on both sides, we get,

$P\{X \leq x'\} = P\{X \leq x\} + P\{x < X \leq x'\}$

$\Rightarrow F(x') - F(x) = P\{x < X \leq x'\} \geq 0$

$\Rightarrow F(x') - F(x) \geq 0$

$\Rightarrow F(x') \geq F(x)$

i.e., F is monotonically non-decreasing in x $\rightarrow$ (1)

Consider $\{x_n\} \downarrow x$

$\Rightarrow \{x < X \leq x_n\} \rightarrow \varphi$

$\Rightarrow P\{x < X \leq x_n\} \rightarrow P(\phi) = 0$

$\Rightarrow F(x_n) - F(x) \rightarrow 0$ as $x_n \rightarrow x$

$\because$ This is true for every sequence $\{x_n\}$. F is continuous from right

Define $A_n = \{X \leq n\}; n = 1, 2...$

Then $A_{n+1} \geq A_n$ and $A_n \uparrow \{X < \infty\} = X^{-1}(R) = \Omega$

Then $P(A_n) \rightarrow P(\Omega) = 1$ as $n \rightarrow \infty$

Hence, $F(+\infty) = 1 \rightarrow (3)$

Define $B_n = \{X \leq -n\}; n = 1, 2, ..$

Now $B_{n+1} \leq B_n$ and

$B_n \downarrow \{X < -\infty\} = \phi$

Then $F(-\infty) = 0 \rightarrow (4)$

Thus from (1) to (4) the define F of a random variable X is non-decreasing continuous on R and $F(-\infty) = 0$ and $F(+\infty) = 1$

ii. Denote by A_n, y the measurable set $[-\infty < X \leq -n, -\infty < Y \leq y]$

When n is a positive integer. For fixed y, the sequence $\{A_n, y\}$ is a contracting sequence whose limit is φ it follows that

$\lim F(-n, y) = \lim P(A_n, y) = P(\lim A_n, y) = P(\phi) = 0$ i.e., $\mathrm{F}(-\infty, \mathrm{y}) = 0$

In similarly way, we have $\mathrm{F}(x - \infty) = 0$

Consider the set $A_n = [-\infty < X \le n, -\infty < Y \le n]$ for positive integer n. Now $\{A_n\}$ is an expanding sequence of measurable sets whose limit is Ω

Hence $\lim F(n, h) = \lim P(A_n) = P(\lim A_n) = P(\Omega) = 1$ i.e., $\mathrm{F}(+\infty, +\infty) = 1$

iii. Note that $D_2 F(x, y) = F(x+h, y+k) - F(x, y+k) - F(x+h, y) + F(x, y)$

$$= P(-\infty < X \le x+h, -\infty < Y \le y+k) - P(-\infty < X \le x, -\infty < Y \le y+k)$$
$$- P(-\infty < X \le x+h, -\infty < Y \le y) + P(-\infty < X \le x, -\infty < Y \le y)$$

Also note that the probability for the rectangle $\{(X, Y)/x < X \le x+h, y < Y \le y+k\}$ which necessary belongs to β^2 equal the expression on the right hand side

Hence $F(x_2, y_2) - F(x_2, y_1) + F(x_1, y_1) - F(x_2, y_1) \ge 0$ holds.

4.4 RANDOM VECTOR:

The vector $\underset{\sim}{X} = (X_1, X_2, ..., X_n)$ defined on (Ω, S, P) into R^n by

$$\underset{\sim}{X}(\omega) = \begin{bmatrix} X_1(\omega) \\ X_2(\omega) \\ . \\ . \\ . \\ X_n(\omega) \end{bmatrix} \forall \omega \in \Omega .$$

Is called an n-dimensional random variable (or) a random vector of size n. If the inverse image of every n-dimensional interval.

$I = \{(x_1, x_2, ... x_n) / -\infty < x_i < x_i a_i \in R; i =, 2, ..., n\}$ is also n, S i.e., if $X^{-1}(I) = \{\omega \in \Omega / X_1(\omega) \le a_i\} \in S, a_i \in R$.

Distributions function of a Random vector:

The function $F()$ is defined of

$F(X_1, X_2, ..., X_n) = P(X_1 \le X_1, X_2 \le x_2, ..., X_n \le x_n) \forall (X_1, x_2, ..., x_n) \in R^n$ is known as a distribution function of random vector.

4.5 INDEPENDENT RANDOM VARIABLES:

We Recall that the joint distribution of a multiple random variable uniquely determines the marginal distribution of the component random variables but in general knowledge of marginal distribution is not enough to determine the joint distribution. Indeed it is quite possible to have an infinite collection of joint densities f_α with given marginal densities.

Definition-1:-we say that X and Y are independent if and only if

i. $F(x,y) = F_1(x)F_2(y)$ for all $(x,y)\in\Re^2$

Lemma:-If X and Y are independent and a<c b<d are real numbers then

ii. $P\{a<X\le c, b<Y\le d\} = P\{a<X\le c\}P\{b<Y\le d\}$

Theorem-1:-

A necessary and sufficient condition for RV's X,Y of the discrete type to be independent is that

$P\{X=x_i, Y=y_j\} = P\{X=x_i\}P\{Y=y_j\}$ for all pairs(x_i, y_j)

Proof:-

Let X,Y be independent. Then from Lemma 1 letting $a\to c$ and $b\to d$ we get

$P\{X=c, Y=d\} = P\{X=c\}P\{Y=d\}$

Conversely $F(x,y) = \sum_B P\{X=x_i, Y=y_i\}$

Where $B = \{(i,j): x_i \le x_i, y_j \le y_i\}$

Then $F(x,y) = \sum_B P\{X=x_i\}P\{Y=y_i\}$

$$= \sum_{x_i\le x}\left[\sum_{y_j\le y} P\{Y=y_i\}\right]P\{X=x_i\} = F(x)F(y)$$

Theorem-2:-Let X and Y be independent RV's and f and g be Boreal-measurable functions. Then f(X) and f(Y) are also independent

Proof:-

We have

$$P\{f(x)\le x, g(y)\le y\} = P\{X\in f^{-1}[-\infty,x], Y\in g^{-1}[-\infty,y]\}$$
$$= P\{X\in f^{-1}(-\infty,x]\}P\{Y\in(-\infty,y]\}$$
$$= P\{f(x)\le x\}P\{g(y)\le y\}$$

Note that a degenerate RV is independent of any RV

Example:-Let X and Y be jointly distributed with PDF

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

Then X and Y are not independent since $f_1(x)=\frac{1}{2}|x|<1$ and $f_2(y)=\frac{1}{2}|y|<1$ are the marginal densities of X and Y respectively. However the RV's X^2 and Y^2 are independent. Indeed

$$P\{X^2\le u, Y^2\le U\}=\int_{-U^{1/2}}^{U^{1/2}}\int_{-u^{1/2}}^{u^{1/2}} f(x,y)dxdy$$

$$=\frac{1}{4}\int_{-U^{1/2}}^{U^{1/2}}\left[\int_{-u^{1/2}}^{u^{1/2}}(1+xy)dx\right]dy$$

$$=u^{1/2}U^{1/2}$$

$$=P\{X^2\le u\}P\{Y^2\le U\}$$

Note that $\phi(x^2)$ and $\phi(Y^2)$ are independent where and are Boreal measurable functions. But X is not a Boreal measurable function of X^2.

Defition-2:- A collection of jointly distributed RV's $X_1, X_2, ...X_n$ is said to be mutually (or) completely independent if and only if $F(x_1,x_2,...,x_n)=\prod_{i=1}^{n}F_i(x_i)$ for all $(x_1,x_2,...,x_n)\in\Re^n$

Where F is the joint DF of $(X_1, X_2, ...X_n)$ and $F_i(i=1,2,....n)$ is the marginal DF of X_i $X_1, X_2, ...X_n$ are said to be pair wise independent if and only if every pair of them are independent.

It is clear that an along of theorem 1 holds but we leave it to the reader to construct it.

Example-:-In example 1 we cannot write $f_\alpha(x_1,x_2,x_3)=f_1(x_1)f_2(x_2)f_3(x_3)$

Except when $\alpha=0$ .It follows that X_1, X_2 *and* X_3 are not independent except when $\alpha=0$.

Theorem-3:-If $X_1, X_2, ...X_n$ are independent every sub collection $X_{i1}, X_{i2}, ...X_{ik}$ of $X_1, X_2, ...X_n$ is also independent.

Remark-1:-It is quite possible for RV's $X_1, X_2, ...X_n$ to be pair wise independent. Without being mutually independent let (X,Y,Z) have the joint PMF defined by $P\{X=x, Y=y, Z=z\}=\begin{cases}\frac{13}{16} if\ (x,y,z)\in\{(0,0,0),(0,1,1),(1,0,1),(1,1,0)\}\end{cases}$

$=\frac{1}{16} if\ (x,y,z)\in\{(0,0,1),(0,1,0),(1,0,0),(1,1,1)\}$

Clearly X,Y,Z are not independent (why?) we have

$$P\{X = x, Y = y\} = \frac{1}{4}, (x, y) \in \{(0,0),(0,1),(1,0),(1,1)\}$$

$$P\{Y = y, Z = z\} = \frac{1}{4}, (x, y) \in \{(0,0),(0,1),(1,0),(1,1)\}$$

$$P\{X = x, Z = z\} = \frac{1}{4}, (x, z) P\{X = x, Y = y\} = \frac{1}{4}, (x, y) \in \{(0,0),(0,1),(1,0),(1,1)\}$$

$$P\{Y = y, Z = z\} = \frac{1}{4}, (y, z) \in \{(0,0),(0,1),(1,0),(1,1)\} \in \{(0,0),(0,1),(1,0),(1,1)\}$$

$$P\{X = x\} = \frac{1}{2}, x = 0, x = 1$$

$$P\{Y = y\} = \frac{1}{2}, y = 0, y = 1$$

$$and P\{Z = z\} = \frac{1}{2}, z = 0, z = 1$$

It follows X and Y and Z and X and Z are pair wise independent.

Definition-3:- A sequence $\{X_n\}$ of RV's is said to be independent if for n=2,3,4…the RV's $X_{1,} X_{2,} \dots X_n$ are independent.

Similarly one can speak of an independent family RV's.

Definition-4:-we say that RV's X and Y are identically distributed if X and Y have the same DF that is

$F_X(x) = F_X(x)$ for all $x \in \Re$

Where F_X *and* F_Y are the DF's of X and Y respectively.

Definition-5:- We say that $\{X_n\}$ is a sequence of independent identically distributed RV's with common law L(X) if $\{X_n\}$ is an independent sequence of RV's and the distribution of $\{X_n\}$, n(1,2,………,n) is the same as that of X.

According to definition 4 X and Y are identically distributed if and only if they have the same distribution. It does not follow that X=Y with probability 1. If $P\{X = Y\} = 1$ we say that X and Y are equivalent RV's. All definition 4 says is that X and Y are identically distributed if and only if $P\{X \in A\} = P\{Y \in A\}$ for all $A \in B$

Nothing is said about the equality of events $P\{X \in A\} and P\{Y \in A\}$.

Definition-6:-Two multiple RV's $(X_1, X_2, \dots X_m)$ and $(Y_1, Y_2, \dots Y_n)$ are said to be independent if $F(x_1, x_2, \dots, x_m, y_1, y_2, \dots, y_n) = F_1(x_1, x_2, \dots, x_m) F_2(y_1, y_2, \dots, y_n)$ for all $(x_1, x_2, \dots, x_m, y_1, y_2, \dots, y_n) \in \Re_{m+n}, where, F, F_1, F_2$ are the joint distribution functions of $(x_1, x_2, \dots, x_m, y_1, y_2, \dots, y_n)$ $(x_1, x_2, \dots, x_m, y_1, y_2, \dots, y_n), (X_1, X_2, \dots, X_m) and (Y_1, Y_2, \dots, Y_n)$ and respectively.

Of course the independence of and does not imply the independence of components $X_1, X_2, ..., X_m$ of X (or) components $Y_1, Y_2, ..., Y_n$ of Y.

Theorem-4:-Let $X = (x_1, x_2, ..., x_m)$ and $(Y_1, Y_2, ..., Y_n)$ be independent RV's. Then the component X_j of $X(j = 1, 2, ...m)$ and the component Y_k of(Y=1,2,….n) are independent RV's. If h and g are Boreal-measurable functions $h(X_1, X_2, ..., X_m)$ and $g(Y_1, Y_2, ..., Y_n)$ are independent.

Remark-2:-It is possible that an RV may be independent of Y and also of Z but X may not be independent of the random vector(Y,Z). See the example in Remark 1.

Let $X_1, X_2, ..., X_n$ be independent and identically distributed RV's with common DF function. Then the DF G of ($X_1, X_2, ..., X_n$) is

Given by

G ($X_1, X_2, ..., X_n$)=$\prod_{j=1}^{n} F(x_j)$

We not for any of the n! permutations $(x_{i1}, x_{i2}, ...x_{in})$ of($X_1, X_2, ..., X_n$) $d = (x_{i1}, x_{i2}, ...x_{in})$ where X $\underline{\underline{d}}$ Y means that X and Y are identically distributed.

Definition-7:-the RV's $X_1, X_2, ..., X_n$ are said to be exchangeable if

($X_1, X_2, ..., X_n$) $d = (x_{i1}, x_{i2}, ...x_{in})$

For all n permutations (i1,i2,….in) of(1,2,…n). The RV's in the sequence $\{X_n\}$ are said to be exchangeable $X_1, X_2, ..., X_n$ are exchangeable for each n.

Clearly if $X_1, X_2, ..., X_n$ are exchangeable then X_i are identically distributed but not necessarily independent.

Example-5:-Suppose that X,Y,Z have joint PDF

$$f(x, y, z) = \begin{cases} \frac{2}{3}(x + y + z), 0 < x < 1, 0 < y < 1, 0 < z < 1 \\ \text{o} \quad \text{,otherwise} \end{cases}$$

Then X,Y,Z are exchangeable but not independent.

Example- Let $X_1, X_2, ..., X_n$ be IID RV's. Let $S_n = \sum_{j=1}^{n} X_j, n = 1, 2, ...$ and $Y_k = X_k - {}^{S_n}\!/_n, k = 1, 2, ...n - 1$. Then $Y_1, Y_2, ..., Y_{n-1}$ are exchangeable

Theorem-5:-Let X,Y be exchangeable RV's . Then X-Y has a symmetric distribution.

The proof is simple.

Theorem-6:-for $e > 0$

(a) $P\left\{\left|X^{s}\right|>e\right\} \leq 2P\left\{X>e/2\right\}$

(b)If $a \geq 0$ such that $P\{X \geq a\} \leq 1-p$ and $P\{X \leq -a\} \leq 1-p$

then $P\left\{\left|X^{s}\right| \geq e\right\} \geq P\{|X|>a+e\}$

4.6 JOINT AND MARGINAL DISTRIBUTION:

The ideas and definitions used in the bivariate case can be readily extended to the multivariate.

The joint distribution of the variables $(X_1, X_2, ..., X_n)$ may be either in terms of the induced probability space say $\left(R_p, B_p, P_{x_1, x_2, ..., x_p}\right)$ or in terms of their distribution function f, defined by

$$f(x_1, x_2, \ldots, x_p) = p\{X_1 \leq x_1, X_2 \leq x_2, \ldots, X_p \leq x_p\}$$

The distribution function has properties analogues to the properties

For a bivariate distribution function conversely any function possessing these properties may be regarded as the distribution function of P jointly distributed random variable.

Here one may think of $2^p - 2$ marginal distributions, of which $[p/1]$ are univariate $[p/2]$ are bivariate… $[p/p-1]$ are (p-1) variate. The marginal distribution function of any $m(<p)$ variables say $X_1, X_2, \ldots, X_m$ will be obtained for X_i ;(i=m+1,m+2,...,p). This may be written as $G_{X_1, X_2 \ldots X_n,}$ and is defined by

$$G(x_1, x_2, \ldots, x_m) = F(x_1, x_2, \ldots, x_m, +\infty, +\infty, \ldots, +\infty)$$

4.7 CONCLUSION:

The study of **random variables** and their associated distributions is fundamental to probability theory and statistics. These concepts provide the mathematical framework for analyzing uncertain outcomes in various applications.

a) **Random Variables and Probability Distributions**: A **random variable** is a function that assigns numerical values to outcomes in a sample space. Its probability distribution describes how probabilities are assigned to different values, distinguishing between **discrete** and **continuous** cases.
b) **Distribution Function**: The **cumulative distribution function (CDF)** describes the probability that a random variable takes on a value less than or equal to a given number. The nature of this function differs for **discrete** and **continuous** random variables.

c) **Vector of Random Variables**: When dealing with multiple random variables simultaneously, they are treated as a **random vector**, which is essential in multivariate statistics and applications like machine learning and economics.
d) **Statistical Independence**: Understanding when multiple random variables are **independent** allows simplification in modeling joint distributions, particularly in probabilistic models and real-world applications.
e) **Joint and Marginal Distributions**: The **joint distribution** describes how multiple random variables behave together, while **marginal distributions** consider individual random variables separately. These concepts are key to understanding dependence structures in probability models.

4.8 SELF ASSESSMENT QUESTIONS:

1. Define a random variable and give an example.
2. What is the probability distribution of a random variable? Explain briefly.
3. Differentiate between discrete and continuous random variables with examples.
4. Define the cumulative distribution function (CDF) and state its properties.
5. What is a random vector? Give an example of its application.
6. Explain statistical independence of two random variables.
7. Differentiate between joint and marginal distributions.
8. Consider a discrete random variable X representing the number of heads in three tosses of a fair coin. Find its probability distribution.
9. Define and derive the properties of the cumulative distribution function (CDF) for both discrete and continuous random variables.

4.9 SUGGESTED READING BOOKS:

1) Modern probability theory by B. R. Bhat, Wiley Eastern Limited.
2) An introduction to probability theory and mathematical statistics by V. K. Rohatgi, John Wiley.
3) An Outline of statistics theory-1, by A.M.GOON, M.K. Gupta and B. Das gupta, the World Press Private Limited, Calcutta.
4) The Theory of Probability by B.V. Gnedenko, MIR Publishers, Moscow.
5) Discrete distributions -N.L. Johnson and S. Kotz, John wiley & Sons.
6) ContinuousUnivariatedistributions,vol.1&2N.L.JohnsonandS.Kotz,JohnWiley&Sons.
7) Mathematical Statistics-Parimal Mukopadhyay, New Central Book Agency (P) Ltd., Calcutta.

Dr. Syed Jilani

LESSON -5

CONDITIONAL DISTRIBUTION AND CHARACTERISTIC FUNCTION

OBJECTIVES:

After studying this unit, you should be able to:

- To understanding the sets and classes of events
- To know the concept of Structure and sets and classes of events
- To acquire knowledge about significance of sets and classes of events
- To understand the purpose and objectives of pivotal provisions of the sets and classes of events

STRUCTURE:

5.1 INTRODUCTION:

In probability and statistics, a **conditional distribution** describes the probability distribution of a random variable given that another variable is known to have a certain value. It helps us understand how one variable behaves under specific conditions.

5.2 CONDITIONAL DISTRIBUTION:

For two-dimensional random variables (x, y) .The joint distribution function $F_{xy}(x,y)$ for any real numbers x and y is given by

$$F_{xy}(x,y) = p(X \le x,\ Y \le y)$$

The conditional distribution function $F_{Y/X}(y/x)$ denotes the distribution function of y when x has already assumed the particular value x.

$$p_{Y/X}(y/x) = p(Y \le y/X=x) = P(A/X=x)$$

5.3 MOMENTS OF CONDITIONAL DISTRIBUTION:

Let us consider, with no loss of generality, the conditional distribution of Y given $X = x$. Since this is a univariable distribution, its moments are to be defined, in terms of either the distribution function $F_Y(./x)$ or the P.D.F (or P.M.F) $f_Y(./x)$. We shall be particularly interested in the conditional mean and the conditional variance.

The conditional mean, if it exists is

$$E\left({}^{Y}/_{x}\right) = \int_{-\infty}^{\infty} y\,dF_Y(y/x)$$

And the conditional variance, if it exists, is

$$var\left({}^{Y}/_{x}\right) = \int_{-\infty}^{\infty} \left[y - E\left({}^{Y}/_{x}\right)\right]^2 dF_Y(y/x)$$

$$= E[(Y-\theta_x)^2/x]$$

Where $\theta_x = E\left({}^{Y}/_{x}\right)$

We have the following theorems:

THEOREM 1:

If X and Y are independent, then $E\left({}^{Y}/_{x}\right) = E(Y)$ and $var\left({}^{Y}/_{x}\right) = var(Y)$

Proof:

$$E\left({}^{Y}/_{x}\right) = \int_{-\infty}^{\infty} Y\,dF_Y(y/x)$$

$$= \int_{-\infty}^{\infty} y\,dH(y), \text{ since } F_Y(y/x) = H(y)$$

$$= E(Y)$$

Also,

$$var\left({}^{Y}/_{x}\right) = \int_{-\infty}^{\infty} [Y-\theta_x]^2 dF_Y(y/x), \text{ Where } \theta_x = E\left({}^{Y}/_{x}\right)$$

$$= \int_{-\infty}^{\infty} [y - E(Y)]^2 dH(y), \text{ from the first result}$$

$$= var(Y)$$

THEOREM 2:

If $E(Y/X)$ exist for almost all values X (i,e.) for all values of X with positives probability or probability –densities then

$$E(Y) = E(Y/X)$$

Proof:

$$E(Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y\,df(x,y)$$

$$= \int_A \left[\int_{-\infty}^{\infty} y\, dF_y({}^{y}/_{x})\right] dG(x)$$

Where A is the set of all x with positive probability or positive probability – densities

$$= \int_A \theta_x\, dG(x)$$

$$= E(\theta_x) = E(y/x)$$

THEOREM-3:

If $E(y/x)$ and $var(y/x)$ exists for almost all values of X_1 then

$$var\,(Y) = E[var(y/x)] + var[E(y/x)]$$

Proof:

$$var\,(Y) = E[Y - \mu_Y]^2$$

$$= EE[(Y - \mu_Y)^2/X]$$

Also,

$$E[(Y - \mu_Y)^2/X] = E([(Y - \theta_x) + (\theta_X - \mu_Y)]^2/X)$$

$$- E[(Y - \theta_x)^2/X]2E[(Y - \theta_X)/X](\theta_X - \mu_Y) + (\theta_X - \mu_Y)^2$$

$$- E[(Y - 0_x)^2/X] + (0_X - \mu_Y)^2$$

Since

$$E[(Y - \theta_X)/X] = \theta_X - \theta_X = 0$$

Thus

$$var(Y) = E\left[var({}^{Y}/_{X}) + (\theta_X - \mu_Y)^2\right]$$

$$= E\left[var({}^{Y}/_{X}) + var\left[E({}^{Y}/_{X})\right]\right]$$

5.4 MATHEMATICAL EXPECTATION:

If 'X' is a discrete random variable which can assume any of the values $x_1, x_2, \ldots, x_n$ with respective probabilities $P_i = P(X = x_i); i = 1, 2, \ldots, n$ then its mathematical expectation is defined as

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

On the other hand if X can take any one of the values $x_i; i = 1, 2 \ldots, \infty$ with respective probabilities P_i then

$$E(X) = \sum_{i=1}^{\infty} P_i x_i; \sum_{i=1}^{\infty} P_i = 1$$

Provided the series is absolutely convergent i.e., provided $\sum_{i=1}^{\infty} |P_i x_i| = \sum_{i=1}^{\infty} P_i |x_i| < \infty$.

If 'X' is a continuous random variable with probability density function $f_X(x)$ then

$$E(X) = \int_{-\infty}^{\infty} Xf_X(x)dx = \int_{-\infty}^{\infty} xdF(x)$$

Provided the integral is absolutely convergent in other words E(X) defined above exist only if

$$\int_{-\infty}^{\infty} |x| f(x)dx < \infty.$$

5.5 CONDITIONAL EXPECTATION:

Definition: Let X and Y be random variables defined on a probability space (Ω,s,ρ) and let 'h' be a borel measurable function. Then the conditional expectation of h(x) given Y written as E [h(x)/y], is a random variable that takes the values E[h(x)/y], defined as

If (X, Y) is Discrete type:

$$E[h(x)/y] = \sum_x h(x).P(X = x \ / Y = y)$$

$$Where\ P(X = x / Y = y) = P(X = x \cap Y = y) / P(Y = y)$$

If (X,Y) is Continuous type:

$$E\left(h(x)/Y\right) = \int_{\infty}^{-\infty} h(x).f_{x/y}(x,y)dx$$

$$\text{Where } f_{X/Y}(x/y) = \frac{f(x,y)}{f_2(y)} \text{ with } f_2(y) > 0$$

F(X,Y) is the joint probability density function of X and Y and $f_2(y)$ the marginal probability density function of Y.

A similar definition may be given for the conditional expectation E[h(y)/x].

NOTE: Expectation of borel-measurable function of a random variable 'x' is always a constant.

But, where as a generally conditional expectation E[h(x)/y] is not a constant but is a random variable.

5.5.1 Properties of conditional expectation:

1. E[c/y] = c, for any constant c.
2. $E[(a_1 g_1(x)+a_2 g_2(x)/y]=a_1.E[g_1(x)/y]+a_2.E[g_2(x)/y]$

   For any Borel-function g1and g2.

3. P(X $\geq$0) =1= E(X/Y) $\geq$0
4. P(X1 $\geq$X2) =1=E(x1/y)>E(x2/y)

5. If X and Y are independent random variables, then

E[X/Y] = E(X) and E[Y/X] = E(Y)
We also have,
V(X/Y) = E(X^2/Y)-[E(X/Y)]^2

$$f(x,y) = e^{-y}, o < x < y < \infty$$

$$f_1(x) = \int_y f(x,y)dy = \int_x^\infty e^{-y}dy = \left[-e^{-y}\right]_x^\infty = -e^{-\infty} + e^{-x} \quad = e^{-x}$$

$$f_2(y) = \int_x f(x,y)\,dx = \int_0^y e^{-y}dx = e^{-y}\int_0^y dx = e^{-y}\left[x\right]_0^y = ye^{-y}$$

$$E(x/y) = \int_x x\,f(x,y)/f_2(y)\,dx$$

$$= \int_x e^{-y}/ye^{-y}dx$$

$$= \int_0^y x/y\,dx$$

$$= 1/y\int_0^y x\,dx$$

$$= 1/y\left[x^2/2\right]_0^y$$

$$= 1/y.y^2/2$$

$$= y/2$$

$$E(y/x) = \int_y y\,f(x,y)/f_1(x)\,dy$$

$$= \int_x^\infty y\,e^{-y}/e^{-x}\,dy$$

$$= \frac{1}{e^{-x}}\int_x^\infty y\,e^{-y}\,dy$$

$$= \frac{1}{e^{-x}}\left[-ye^{-y} - e^{-y}\right]_x^\infty$$

$$= \frac{1}{e^{-x}}\left[-e^{-y}(y+1)\right]_x^\infty$$

$$= \frac{1}{e^{-x}}\left[-e^{-\infty}(\infty+1) + -e^{-x}(x+1)\right]$$

$$= \frac{1}{e^{-x}}\left[e^{-x}(x+1)\right]$$

$$E(y/x) = x+1$$

5.6 CHARACTERISTIC FUNCTION:

Definition: Characteristic function is defined as

$$\phi_X(t) = E\left(e^{itX}\right)$$

$$= \int e^{itx} f(x)dx \text{ (for continuous probabilty distribution)}$$

$$= \sum_x e^{itx} P(x) \text{ (for discrete probability distribution)}$$

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

$$\phi_X(t) = \int_{-\infty}^{\infty} e^{itx} F(x).$$

Obviously $\phi(t)$ is a complex valued function of real variable ' t ' and it may be noted that

$$\left|\phi_X(t)\right| = \left|E(e^{itX})\right| \le E\left(\left|e^{itX}\right|\right) = \int \left|e^{itx}\right| f(x)dx = \int f(x)dx = 1$$

Since $\left|e^{itx}\right| = \left|\cos tx + i \text{ sintx}\right|^{1/2} = \left(\cos^2 tx + \sin^2 tx\right)^{1/2} = 1$.

Since $\left|\phi_X(t)\right| \le 1$ characteristic function $\phi_X(t)$ always exists though $M_X(t)$ may not exist. This is the striking advantage of characteristic function over m.g.f.

Another Definition: Let X be a random variable. Then the complex valued function φ defined on R by

$$\phi_X(t) = E\left(e^{itX}\right) = E\left(\cos tX + i\sin tX\right) = E\left(\cos tX\right) + iE\left(\sin tX\right)$$

Where $i = \sqrt{-1}$ is the imaginary unit is called as the characteristic function (CF) of the random variable X.

On the discrete case CF is

$$\phi(\mathrm{t}) = \sum_{\mathrm{k}} \left(\text{cost}\, x_{\mathrm{k}} + \text{isint}\, x_{\mathrm{k}}\right) \mathrm{P}\left(\mathrm{X} = x_{\mathrm{k}}\right)$$

And in the continuous case CF is $\phi(\mathrm{t}) = \int_{-\infty}^{\infty} \text{cost}\, x f(x)dx + i\int_{-\infty}^{\infty} \text{sint}x f(x)dx$

<u>**Note**</u>: the study of characteristic function requires the knowledge of complex analysis in particular that of complex intersection.

<u>**Note**</u>: CF is also called as the fourier transform of $F(x)$ e^{itx} is called the kernel of the transform.

5.6.1 Some Simple Properties of Characteristic function:

If $\phi(t)$ is the Characteristic function (CF) of general distribution function $F(x)$ then

1. $\phi(.)$ is continuous.

2. $|\phi(t)| \le \phi(0) = F(+\infty) - F(-\infty) = 1$.
3. $\phi(-t) = \bar{\phi}(t)$.

where $\phi(t)$ is C.F. of a random variable X. Then the C.F. of a+bx is $e^{ita}\phi(bt)$. In particular $\phi(.)$ is the C.F of –X $\phi(.)$ is real X is symmetric about origin.
4. If $\phi_n(t)$ is an C.F of random variable then $\sum r_n\phi_n(t)$ is C.F of R.V's $r_n \ge 0, \sum r_n = 1$.

Proof:

1. For every t' is neighbourhood of t, we have

$$|\phi(t') - \phi(t)| = \left|E\left(e^{it'X}\right) - E\left(e^{itX}\right)\right| = \left|E\left(e^{it'X} - e^{itX}\right)\right| \le E\left|e^{it'X} - e^{itX}\right|$$

($\because$ *for any complex* rv Z, *we have* $|E(Z)| \le E(|Z|)$)

$$\Rightarrow |\phi(t') - \phi(t)| \le \int_{-\infty}^{\infty} \left|e^{it'x} - e^{itx}\right| f(x)dx$$

But $\left|e^{it'x} - e^{itx}\right| = \left|e^{it'x}\left(1 - e^{i(t-t')x}\right)\right| = \left|e^{it'x}\right|\left|1 - e^{i(t-t')x}\right| \to 0$ as $t \to t'$ (1)

Further, $\left|e^{it'x} - e^{itx}\right| \le \left|e^{it'x}\right| + \left|-e^{itx}\right| = \left|e^{it'x}\right| + \left|e^{itx}\right| = 1 + 1 = 2$ (2)

$$\left[\text{since} \quad \left|e^{itx}\right| = |\cos tx + i\ \text{sintx}|^{1/2} = (\cos^2 tx + \sin^2 tx)^{1/2} = 1\right]$$

In view of (1) and (2), using dominated convergence theorem, we see that

$$|\phi(t') - \phi(t)| \to 0 \text{ as } |t - t'| \to 0 \quad \Rightarrow \quad \phi(t) \text{ is continous}$$

Hence the proof.

2. For any C.F. $\phi_X(t)$, we have

$$|\phi_X(t)| = \left|E\left(e^{itX}\right)\right| \le E\left(\left|e^{itX}\right|\right) = \int_{-\infty}^{\infty} \left|e^{itx}\right| f(x)dx$$

But, we have $\left|e^{itx}\right| = |\cos tx + i\ \text{sintx}|^{1/2} = (\cos^2 tx + \sin^2 tx)^{1/2} = 1$.

Therefore $|\phi_X(t)| \le \int_{-\infty}^{\infty} f(x)dx = 1$ (1)

Also we have $\phi(0) = \int_{-\infty}^{\infty} e^0 f(x)dx = \int_{-\infty}^{\infty} f(x)dx = F(\infty) - F(-\infty) = 1$ (2)

$$[\text{ since, we have } \int_a^b f(x)dx = F(b) - F(a)\ ]$$

Thus, from (1) and (2), we have $|\phi(t)| \le \phi(0) = 1$

which implies that $\phi(t)$ is bounded by $\phi(0) = 1$.

3. We have

$$\phi_X(-t) = E\left(e^{-itX}\right) = E\left(\cos(-tX) + i\sin(-tX)\right) = E(\cos tX - i\sin tX)$$

$$= E(\cos tX) - i\,E(\sin tX) = \overline{\phi_X}(t)$$

Thus $\phi_X(-t) = \overline{\phi_X}(t)$.

4. CF of $a + bX$ is

$$E\left(e^{it(a+bX)}\right) - \int_{-\infty}^{\infty} e^{it(a+bx)} f(x)dx \; - e^{ita}\int_{-\infty}^{\infty} e^{itbx} f(x)dx = e^{ita}\phi(tb)$$

In particular, if we take a=0 and b=-1, then the ch. function of

$Y = a + bX = -X$ is given as

$$\phi_Y(t) = \phi_{-X}(t) = E\left[e^{-itX}\right] = \phi_X(\;\;t) = \bar{\phi}_X(t)$$

Thus if $\phi_X(t)$ is Ch. Fn. of X, then $\bar{\phi}_X(t)$ is Ch. Fn. if $-X$

If X is symmetric about origin then X and $-X$ will have the same distribution function as explained below

$$F(x) = P(X \le x) = P(X \ge -x)$$

$$= P(-X \le x)$$

$$= G(x)$$

Where F and G are respectively the DFs of X and $\;\;X$.

Thus, X and $-X$ will have the same DF and have the same Ch. function therefore $\phi(t) = \bar{\phi}(t)$

$$\Rightarrow \int_{-\infty}^{\infty} \cos tX\, f(x)dx + i\int_{-\infty}^{\infty} \sin tX\, f(x)dx$$

$$= \int_{-\infty}^{\infty} \cos tX\, f(x)dx - i\int_{-\infty}^{\infty} \sin tX\, f(x)dx$$

$\Leftrightarrow \int_{-\infty}^{\infty} \sin tX = 0$ is the imaginary part varnishes.

Thus $\phi(t)$ is real

Conversely if $\varphi(t) = \bar{\phi}(t)$, then X and –X will have the same CHF and hence the same DF (since the CHF determines uniquely the DF) thus X is symmetric.

The moment problem: We have seen that moments may not exists, but a characteristic function will always exist. However, if moments of all order exist, then moments can determine the characteristic function. If the series $\sum_{r=0}^{\infty} \frac{\mu_r' e^r}{r!}$ converges for a certain P>0.

This will be so it $\sum_n \frac{V_n e^n}{n!}$ is convergent.

5.6.2 Characteristic Functions and Moments:

Derivatives of the characteristics function $\phi(\mathrm{t})$:

We have by definition, $\phi(\mathrm{t}) = \phi(\mathrm{t}) = \mathrm{E}\left(\mathrm{e}^{\mathrm{itx}}\right)$ and $\phi(0) = \mathrm{E}\left(\mathrm{e}^{0}\right) = 1$ (1)

We also know that $\phi(\mathrm{t})$ is continuous and the derivative of ϕ at t=0 is given by

$$\phi'(0) = \lim_{h\to 0} \frac{\phi(h)-\phi(0)}{h} = \lim_{h\to 0} \frac{\mathrm{E}\left(\mathrm{e}^{ihX}\right)-1}{\mathrm{h}} \qquad \text{(from(1))}$$

$$= \lim_{h\to 0} \mathrm{E}\left[\frac{\mathrm{e}^{ihX}-1}{h}\right] = \lim_{h\to 0} \int_{-\infty}^{\infty} \left[\frac{\mathrm{e}^{ihX}-1}{h}\right] f(x)\,dx = \int_{-\infty}^{\infty} \lim_{h\to 0} \left[\frac{\mathrm{e}^{ihX}-1}{h}\right] f(x)\,dx \qquad (2)$$

$$\left[\text{By dominated convergence theorem}\right]$$

But we have,

$$e^{ihx} = 1 + ihx + \frac{(ihx)^2}{\lfloor 2} + \frac{(ihx)^3}{\lfloor 3} + \ldots$$

$$\Rightarrow \frac{e^{ihx}-1}{h} = ix + \frac{h(ix)^2}{\lfloor 2} + \frac{h^2(ix)^3}{\lfloor 3} + \ldots.. \Rightarrow \lim_{h\to 0} \frac{e^{ihx}-1}{h} = ix \qquad (3)$$

Upon using equation (3) in equation (2) we get,

$$\phi'(0) = \int_{-\infty}^{\infty} ix.f(x)\,dx = i\int_{-\infty}^{\infty} xf(x)\,dx = i\,E(X) = i\,\mu_1 \quad (\text{say})$$

Thus, if ϕ' exist and is finite at the origin, then the first moment of X i.e. $E(X)$ exists and is finite.

$$\phi(t) = \lim_{h\to 0} \frac{\phi'(t+h)-\phi'(t)}{h} \quad = \lim_{h\to 0} \int ix\left[\frac{e^{i(t+h)x}-e^{itx}}{h}\right]f(x)\,dx$$

$$= \lim_{h\to 0}\int ix\left[\frac{e^{ihx}-1}{h}\right]e^{itx}f(x)\,dx \quad = \int \lim_{h\to 0}\left[\frac{e^{ihx}-1}{h}\right]ix.e^{itx}.f(x)\,dx$$

$$\phi(t) = \sum_{r=0}^{\infty}\frac{(it)^r}{r!}\mu_r^1 \le \int |x|^r.f(x)\,dx \quad \left(\because \left|e^{itx}\right|\le 1\right)$$

$$< \ \infty \quad (\text{from } (2))$$

$$\left(\lim_{h\to 0} \quad \left|\frac{\left(e^{ihx}-1\right)}{h}\right| \le |x|\right)\frac{\left(e^{ihx}-1\right)}{h} = ix \qquad (4)$$

$$\phi^r(0) = i^r\int x^r f(x)\,dx = i^r \ E(X) \ = i^r\mu_r^1$$

Theorem: If the r^{th} absolute moment of a random variable X is finite, then the characteristic function $\phi(\mathrm{t})$ is differentiable ' r ' times and

$$\phi^r(t) = \int (ix)^r e^{itx} f(x)\,dx \qquad (1)$$

Where $f(x)$ is pdf of X and t is any real number.

Proof: we have given rth absolute moment id finite.

$$\therefore \int |x|^r f(x)\,dx < \infty \qquad (2)$$

The first ordered derivation of $\phi(\mathrm{t})$ is given by

$$\phi'(t) = \lim_{h\to 0}\frac{\phi(t+h)-\phi(t)}{h}$$

$$= \lim_{h\to 0}\int\left[\frac{e^{i(t+h)x}-e^{itx}}{h}\right]f(x)\,dx$$

$$= \lim_{h\to 0}\int e^{itx}\left[\frac{e^{ihx}-1}{h}\right]f(x)\,dx$$

$$= \int e^{itx}\lim_{h\to 0}\left[\frac{e^{ihx}-1}{h}\right]f(x)\,dx \qquad (3)$$

$\Big(\lim_{h\to 0}$ can be taken under the integral sign by dominated convergence theorem because,

$$\left.\left|\frac{e^{itx}\left(e^{ihx}-1\right)}{h}\right| \le |x|\right)$$

But we have, $\lim\limits_{h\to 0}\dfrac{\left(e^{ihx}-1\right)}{h}=ix$ (4)

Upon using (4) in (3), we get

$$\phi'(t)=\int ix.e^{itx}f(x)\,dx$$

Similarly, we have

$$\phi'(t)=\lim_{h\to 0}\frac{\phi'(t+h)-\phi'(t)}{h}$$

$$=\lim_{h\to 0}\int ix\left[\frac{\mathrm{e}^{i(t+h)x}-e^{itx}}{h}\right]f(x)\,dx$$

$$=\lim_{h\to 0}\int ix\left[\frac{\mathrm{e}^{ihx}-1}{h}\right]e^{itx}f(x)\,dx$$

$$=\int\lim_{h\to 0}\left[\frac{\mathrm{e}^{ihx}-1}{h}\right]ix.e^{itx}.f(x)\,dx$$

Upon using (4) in (5), we get

$$\phi''(t)=\int(ix)^2 e^{itx}f(x)\,dx$$

Continuing by induction, the rth derivative of $\phi(t)$ is given by

$$\phi^r(t)=\int(ix)^r e^{itx}f(x)\,dx$$

But we have,

$$\left|\phi^r(t)\right|=\left|\int(ix)^r e^{itx}f(x)\,dx\right|$$

$$\leq\int|x|^r.f(x)\,dx \quad \left(\because\left|e^{itx}\right|\leq 1\right)$$

$$< \infty \quad (\text{from }(2))$$

Therefore, all moments of up to order ' r ' would be finite.

Note: From equation (1) we have,

$$\phi^r(0)=i^r\int x^r f(x)\,dx$$

$$=i^r\ E(X)$$

$$=i^r\mu_r^1$$

Note: The moments may not exist, but a characteristic function will always exit. However, if moments of all order exist, then moments can be determined using the characteristic function as given below

$$\phi(t)=\sum_{r=0}^{\infty}\frac{(it)^r}{r!}\mu_r^1$$

1.7 CONCLUSION:

The study of **conditional distributions, expectation, and characteristic functions** provides fundamental tools for understanding probability and statistics. These concepts allow us to analyze dependencies between random variables, derive important statistical properties, and characterize probability distributions effectively.

a) **Conditional Distribution**: This describes the probability distribution of a random variable given that another variable has taken a specific value. It is crucial in Bayesian inference and decision theory.
b) **Moments of Conditional Distribution**: Moments (such as mean and variance) of a conditional distribution provide insights into the expected behavior of a random variable given certain conditions.
c) **Mathematical Expectation**: Expectation (or expected value) is a measure of the central tendency of a random variable. It is used extensively in probability theory, economics, and finance.
d) **Conditional Expectation**: The expected value of a random variable given another variable captures dependencies and is widely used in stochastic processes and statistical modeling.
e) **Characteristic Function**: This function provides a powerful way to describe the distribution of a random variable. It is especially useful in proving limit theorems and deriving moments.
f) **Properties and Moments of the Characteristic Function**: The characteristic function uniquely determines the distribution of a random variable, and its derivatives help compute moments such as mean and variance.

1.8 SELF ASSESSMENT QUESTIONS:

1 Define conditional distribution and explain its significance.
2. Derive the formula for the conditional expectation $E(X \mid Y = y)$.
3. State and explain the properties of conditional expectation.
4. Define the characteristic function of a random variable and list its properties.
5. How can the moments of a random variable be obtained from its characteristic function?
6. Prove the law of iterated expectation.
7. Explain the relationship between independence and conditional expectation

1.9 SUGGESTED READINGS:

1) Modern probability theory by B. R. Bhat, Wiley Eastern Limited.
2) An introduction to probability theory and mathematical statistics by V. K. Rohatgi, John Wiley.
3) An Outline of statistics theory-1, by A.M.GOON, M.K. Gupta and B. Das gupta, the World Press Private Limited, Calcutta.
4) The Theory of Probability by B.V. Gnedenko, MIR Publishers, Moscow.
5) Discrete distributions -N.L. Johnson and S. Kotz, John wiley & Sons.
6) ContinuousUnivariatedistributions,vol.1&2N.L.JohnsonandS.Kotz,JohnWiley&Sons.
7) Mathematical Statistics-Parimal Mukopadhyay, New Central Book Agency (P) Ltd., Calcutta.

Dr. Syed Jilani

LESSON -6
MOMENT'S INEQUALITIES

OBJECTIVES:

After studying this unit, you should be able to:

- To understanding the moment's inequalities
- To know the concept of Structure and moment's inequalities
- To acquire knowledge about significance of moment's inequalities
- To understand the purpose and objectives of pivotal provisions of the moment's inequalities

STRUCTURE:

6.1 INVERSION THEOREM:

Statement: Let $\varphi_x(t)$ is the characteristic function of $F(t)$ and $(a-h, a+h)$ is continuity interval of $F(t)$ then $F(a+h)-F(a-h)=\lim_{T\to\infty}\frac{1}{\pi}\int_{-T}^{T}\frac{\sinh t}{t}e^{-ita}\varphi_X(t)\,dt$

so that if there are two distribution functions, they agree at all continuity points and therefore both are identical.

Proof: To show that

$$F(a+h)-F(a-h)=\lim_{T\to\infty}\frac{1}{\pi}\int_{-T}^{T}\frac{\sinh\ t}{t}e^{-ita}\varphi_X(t)dt$$

R.H.S:

Let $J = \frac{1}{\pi}\int_{-T}^{T}\frac{\sinh t}{t}e^{-ita}\left[\int_{-\infty}^{\infty}e^{itx}F(x)dx\right]dt$

Where $F(x)$ is p.d.f of X . x is continuous

$$\Rightarrow J = \frac{1}{\pi}\int_{-T}^{T}\frac{\sinh t}{t}\left[\int_{-\infty}^{\infty}e^{it(x-a)}F(x)\,dx\right]dt$$

$$\Rightarrow J = \frac{1}{\pi}\int_{-\infty}^{\infty}F(x)\left[\int_{-T}^{T}\frac{\sinh t}{t}e^{it(x-a)}dt\right]dx$$

consider the inner integral

$$\text{let I} = \int_{-T}^{T}\frac{\sinh t}{t}e^{it(x-a)}dt$$

$$\Rightarrow \text{I} = \int_{-T}^{T}\frac{\sinh t}{t}\left[\cos(x-a)t + i\sin(x-a)t\right]dt$$

$$\Rightarrow \text{I} = \int_{-T}^{T}\frac{\sinh t\cos(x-a)t}{t}dt + i\int_{-T}^{T}\frac{\sinh t\sin(x-a)t}{t}dt$$

[since cosx is even function cos(-x)=cosx,sinx is odd function sin(-x)=-sinx]

$$\therefore \quad \text{I} = 2\int_{0}^{T}\frac{\sinh t\cos(x-a)t}{t}dt \neq 0$$

[since $\int_{-\infty}^{\infty}\sin x$ dx =0, $\int_{-\infty}^{\infty}$ cosx dx =2$\int_{0}^{\infty}\cos x$ dx]

$$\Rightarrow \text{I} = \int_{0}^{T}\frac{\sin(h+x-a)t}{t}dt + \int_{0}^{T}\frac{\sin(h-x+a)t}{t}dt \qquad \left[\because 2\sin\text{A}\cos\text{B} = \sin(\text{A}+\text{B}) + \sin(\text{A}-\text{B})\right]$$

$$\Rightarrow \lim_{T\to\infty}\text{I} = \int_{0}^{\infty}\frac{\sin(x+h-a)t}{t}dt + \int_{0}^{\infty}\frac{\sin(h-x+a)t}{t}dt$$

Since from the standard integral of improper integrals.

$$\int_{0}^{\infty}\frac{\sinh t}{t}dt = \frac{\pi}{2} \text{ if h} > 0$$

$$= 0 \text{ if h} = 0$$

$$= -\frac{\pi}{2} \text{ if h} < 0$$

Using the above results we have.

$$\lim_{T\to\infty} \mathrm{I} = \pi \text{ if } x \in (a-h, a+h)$$

$$= 0 \text{ for all other values of } x$$

$$\lim_{T\to\infty} J = \frac{1}{\pi}\int_{-\infty}^{\infty} F(x)\left\{\int_{o}^{\infty}\frac{\sin(x+h-a)t}{t}dt + \int_{0}^{\infty}\frac{\sin(h-x+a)t}{t}dt\right\}dx$$

$$\Rightarrow \lim_{T\to\infty} J = \frac{1}{\pi}\int_{-\infty}^{\infty}\pi F(x)\,dx$$

$$\Rightarrow \lim_{T\to\infty} J = \int_{-\infty}^{a-h} F(x)dx + \int_{a-h}^{a+h} F(x)dx + \int_{a+h}^{\infty} F(x)dx$$

$$\Rightarrow \lim_{T\to\infty} J = \int_{a-h}^{a+h} F(x)dx$$

$$\Rightarrow \lim_{T\to\infty} J = F(a+h) - F(a-h)$$

L.H.S=R.H.S

6.2 CHEBYSHEV'S INEQUALITY:

Statement: If X is any random variable, then $P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}$ where $\mu = E(X)$ and $\sigma^2 = \mathrm{var}(X)$ and $k > 0$

Proof: Chebychev's inequality is also a special case of basic inequality. Therefore, first we have to state and prove the basic inequality .

in basic inequality, if we take

$$h(x) = (X-\mu)^2 \; and \;\; \varepsilon = k^2.\sigma^2$$

We get Cheybychev inequality and is given by

$$P\left\{(X-\mu)^2 \geq k^2.\sigma^2\right\} \leq \frac{E\left\{(X-\mu)^2\right\}}{k^2.\sigma^2} \quad \Rightarrow P\left\{|X-\mu| \geq k\sigma\right\} \leq \frac{1}{k^2}$$

$$\left(\because \sigma^2 = E(X-\mu)^2 = V(X)\right)$$

Hence the proof.

Chebychev's inequality as a special case of Markov inequality.

Statement: If X is a random variable with mean $E(X) = \mu$ and $V(X) = \sigma^2$, then for any k >0, we will have

$$P\left\{|X-\mu| \geq k\sigma\right\} \leq \frac{1}{k^2} \qquad \text{Or} \qquad P\left[\frac{|X-\mu|}{\sigma} \geq k\right] \leq \frac{1}{k^2}$$

Proof: This can be deduced from Markov inequality. Therefore, first we have to prove Markov theorem, then we have to deduce. Chebychev inequality as a special case of Markov inequality.

From Markov inequality, we have

$$P(|Y| \geq c) \leq \frac{E\left[|Y|^r\right]}{c^r} \text{ where } c > 0 \text{ and } r > 0$$

In the above Markov inequality, take $r = 2$, Y=X-μ and c=kσ where k>0 so that c>0

Then we get

$$P\{|X-\mu| \geq k\sigma\} \leq \frac{E\left[(X-\mu)^2\right]}{k^2\sigma^2}$$

But, we have $E\left[(X-\mu)^2\right] = \sigma^2 \quad \therefore P\left[|X-\mu| \geq k\sigma\right] \leq \frac{1}{k^2}$ Or $P\left[|X-\mu|/\sigma \geq k\right] \leq \frac{1}{k^2}$

Convex function :

Let $f(x)$ be a real valued Borel function defined on an interval.

$I(\text{finite or infinite}) \subseteq R$

Now, the function f is said to be convex if for every pair of points x, x' of I.

$$f\left(\frac{x+x'}{2}\right) \leq \frac{f(x)+f(x')}{2}$$

If f is twice- differentiable on I, convexity is equivalent to $f(x'') \geq 0$ on I

A continuous function is always continuous.

An alternative definition of convex function is that, for every $x_0 \in I$, there corresponds a number $\lambda(x_0)$ such that for all $x \in I$.

$$\lambda(x_0)(x-x_0) \leq f(x) - f(x_0)$$

The LHS may be interpreted as the tangent through x_0, if it exists. The above equation implies that all the points of the curve $f(x)$ are above this tangent line. For example $|X|^r$ is the a convex function of x.

6.3 JENSEN'S INEQUALITY:

Statement: If X is a random variable with finite mean $E(X)$ and f(x) is a convex function then show that $\mathrm{E}\left[\mathrm{f(X)}\right] \geq \mathrm{f}[E(X)]$.

Proof: Let X be a random variable whose value lie in $I \subseteq R$. Since f(x) is a convex function, it immediately follows.

$\lambda(x_0)(x-x_0) \le f(x)-f(x_0)$

Where $x_0 \in I$ and $\lambda(x_0)$ is a number which corresponds to x_0.Take x_0=E(X) in the above equation (x). Then we get,

$\lambda(E(X))$ [x-E(X)] $\le f(x)-f(E(X))$.

Taking expectation on both sides, we get

$\lambda(E(X))$ [E(X)-E(X)] $\le E[f(X)]-f(E(X))$

$\Rightarrow E[f(X)]-f(E(X)) \ge 0$. $[\because E(X)-E(X)=0]$

$E[f(X)] \ge f(E(X))$

Hence the proof.

6.4 CAUCHY-SCHWARTZ INEQUALITY.

State and prove Holder's inequality hence obtain Cauchy-Schwartz inequality.

Holder's Inequality : $E\left(|XY|\right) \le \left[E\left(|X|^{\mathrm{r}}\right)\right]^{1/\mathrm{r}} . \left[E\left(|Y|^{s}\right)\right]^{1/s}$ (1)

Where $\mathrm{r}>1$ and $\frac{1}{\mathrm{r}}+\frac{1}{s}=1$.

Proof: Consider $\phi(p)=\frac{p^{\mathrm{r}}}{\mathrm{r}}+\frac{p^{-s}}{s}; p>0$

$\phi(p)$ is minimum at $p=1$ with $\phi(1)=1$. That is $\phi(p)\ge\phi(1)=1;$ for any $p>0$

Let $p=p_0=\frac{a^{1/s}}{b^{1/\mathrm{r}}}$ with $a>0; b>0$

Now, $\phi(p_0)=\frac{1}{\mathrm{r}}.\frac{a^{\mathrm{r}/s}}{b^{\mathrm{r}/\mathrm{r}}}+\frac{1}{s}.\frac{a^{-s/s}}{b^{-s/\mathrm{r}}}$

$$=\mathrm{r}^{-1}.\frac{a^{\mathrm{r}/s}}{b}+s^{-1}.\frac{b^{s/\mathrm{r}}}{a} \ge 1=\phi(1) \quad (2)$$

$$\Rightarrow \frac{\mathrm{r}^{-1}a^{1+\mathrm{r}/s}+s^{-1}b^{1+s/\mathrm{r}}}{ab} \ge 1$$

$$\Rightarrow \mathrm{r}^{-1}a^{(\mathrm{r}+s)/s}+s^{-1}b^{(\mathrm{r}+s)/\mathrm{r}} \ge ab \quad (3)$$

But we have

$$\frac{1}{\mathrm{r}}+\frac{1}{s}=1 \Rightarrow \frac{\mathrm{r}+s}{\mathrm{r}s}=1 \quad (4)$$

Upon using (4) and (3), we get

$$r^{-1}a^{r}+s^{-1}b^{s}\geq ab \qquad \Rightarrow \qquad \frac{a^{r}}{r}+\frac{b^{s}}{s} \geq ab$$

Putting $a=\dfrac{|X(\omega)|}{\left[E|X|^{r}\right]^{1/r}}$; $b=\dfrac{|Y(\omega)|}{\left[E|Y|^{s}\right]^{1/s}}$

We get ; $\dfrac{1}{r}.\dfrac{|X(\omega)|^{r}}{E\left(|X|^{r}\right)}+\dfrac{1}{s}.\dfrac{|Y(\omega)|^{s}}{E\left(|Y|^{s}\right)}\geq\dfrac{|X(\omega)||Y(\omega)|}{\left[E\left(|X|^{r}\right)\right]^{1/r}\left[E\left(|Y|^{s}\right)\right]^{1/s}}$

$$\geq\frac{|X(\omega)||Y(\omega)|}{\left[E\left(|X|^{r}\right)\right]^{1/r}\left[E\left(|Y|^{s}\right)\right]^{1/s}}$$

Taking expectation on both sides

$$\frac{1}{r}.\frac{E\left(|X|^{r}\right)}{E\left(|X|^{r}\right)}+\frac{1}{s}.\frac{E\left(|Y|^{s}\right)}{E\left(|Y|^{s}\right)}\geq\frac{E\left(|XY|\right)}{\left[E\left(|X|^{r}\right)\right]^{1/r}\left[E\left(|Y|^{s}\right)\right]^{1/s}}$$

$$\Rightarrow\frac{1}{r}+\frac{1}{s}\geq\frac{E\left(|XY|\right)}{\left[E\left(|X|^{r}\right)\right]^{1/r}\left[E\left(|Y|^{s}\right)\right]^{1/s}}$$

$$\Rightarrow E\left(|XY|\right)\leq\left[E\left(|X|^{r}\right)\right]^{1/r}\left[E\left(|Y|^{s}\right)\right]^{1/s} \qquad \left(\because \frac{1}{r}+\frac{1}{s}=1\right)$$

Hence, the theorem.

6.5 MINKOWSKI INEQUALITY:

Statement: For $r\geq 1$

$$\left[E\left(|X+Y|^{r}\right)\right]^{1/r}\leq\left[E\left(|X|^{r}\right)\right]^{1/r}+\left[E\left(|Y|^{r}\right)\right]^{1/r}$$

Proof:

$$E\left(|X+Y|^{r}\right)=E\left(|X+Y|\,.\,|X+Y|^{r-1}\right)$$

$$\leq E\left(|X|.|X+Y|^{r-1}\right)+E\left(|Y||X+Y|^{r-1}\right) \qquad (1)$$

But from Holder’s inequality, we have

$$E\left(|X||X+Y|^{r-1}\right)=E\left(\left|(X)(X+Y)^{r-1}\right|\right)$$

$$=\left[E\left(|X|^r\right)\right]^{1/r}\left[E\left(|X+Y|^{s(r-1)}\right)\right]^{1/s}$$

But, we have $\frac{1}{r}+\frac{1}{s}=1$ so that $s=\frac{r}{r-1}$

$$\therefore E\left(|X||X+Y|^{r-1}\right)=\left[E\left(|X|^r\right)\right]^{1/r}\left[E|X+Y|^r\right]^{(1-\frac{1}{r})} \quad (2)$$

Similarly,

$$E\left(|Y||X+Y|^{r-1}\right)=\left[E\left(|Y|^r\right)\right]^{1/r}\left[E|X+Y|^r\right]^{(1-\frac{1}{r})} \quad (3)$$

Substituting (2) & (3) in (1) we get,

$$E\left(|X+Y|^r\right)\le\left[E\left(|X|^r\right)\right]^{1/r}\left(E\left(|X+Y|^r\right)\right)^{(1-\frac{1}{r})}+\left[E\left(|Y|^r\right)\right]^{\frac{1}{r}}\left[E\left(|X+Y|^r\right)\right]^{(1-\frac{1}{r})}$$

$$\Rightarrow \frac{E\left(|X+Y|^r\right)}{\left(E\left(|X+Y|^r\right)\right)^{(1-\frac{1}{r})}}\le\left[E\left(|X|^r\right)\right]^{\frac{1}{r}}+\left[E\left(|Y|^r\right)\right]^{\frac{1}{r}}$$

$$\Rightarrow \frac{E\left(|X+Y|^r\right)}{E\left(|X+Y|^r\right)\left[E|X+Y|^r\right]^{-1/r}}\le\left[E\left(|X|^r\right)\right]^{1/r}+\left[E\left(|Y|^r\right)\right]^{1/r}$$

$$\left[E\left(|X+Y|^r\right)\right]^{1/r}\le\left[E\left(|X|^r\right)\right]^{1/r}+\left[E\left(|Y|^r\right)\right]^{1/r}$$

Hence, the Minkowski Inequality

6.6 MARKOV'S INEQUALITY:

STATEMENT: If X is any random variable, then $P\left\{|X|\ge K\right\}\le\frac{E\left\{|X|^r\right\}}{K^r}$ where r>0 and k>0.

Proof: let 'S' be the subset of real line, i.e., $S\subseteq R$

Let us define,

$S=\left\{x\in R/|x|\ge k\right\}$, where $k>0$

Suppose $f(x)$ is the p.d.f. of random variable X

Consider $|x|\ge k$

$\Rightarrow |x|^r\ge k^r$, where $r>0$

$$\Rightarrow |x|^{r} f(x) \geq k^{r} f(x) \qquad \therefore f(x) > 0$$

$$\Rightarrow \int_{S} |X|^{r} f(x) dx \geq k^{r} \int_{s} f(x) dx \rightarrow (1)$$

But, we have

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

$$= \int_{S} |x|^{r} f(x) dx + \int_{R-S} |x|^{r} f(x) dx$$

Since $\int_{R-S} |x|^{r} f(x) dx \geq 0$, we have

$$E\{|X|^{r}\} \geq k^{r} \int_{s} f(x) dx$$

$$\Rightarrow \int_{S} f(x) dx \leq \frac{E\{|X|^{r}\}}{k^{r}} \rightarrow (2)$$

But we have $S = \{x \in R / |x| \geq k\}$

Therefore,

$$\int_{S} f(x) dx = P\{x \in R / |x| \geq k\}$$

$$= P\{|X| \geq k\} \rightarrow (3)$$

Thus from (2) and (3) we have

$$P\{|X| \geq k\} \leq \frac{E\{|X|^{r}\}}{k^{r}} \qquad \text{(or)} \qquad P\{|X| \geq k\} \leq E\frac{\{|X|^{r}\}}{k^{r}}$$

Hence the proof.

Note: Markov's inequality may also be stated as $P\{|X| > k\} < \frac{E\{|X|^{r}\}}{k^{r}}$

Markov's inequality as a special case of basic inequality.

Statement: let $h(x)$ be a non-negative Borel measurable function of a random variable X. Now ,if $E\{h(x)\}$ exits, then for $\varepsilon > 0$,

$$p\{h(x) \geq \varepsilon\} \leq \frac{E\{h(x)\}}{\varepsilon}$$

Proof: let S be a subset of real line i.e., $S \subseteq R$

Let us define $S = \{x \in R / h(x) \geq \varepsilon\}$

Suppose $f(x)$ is the pdf of X. Now, consider $h(x) \geq \varepsilon$

$$\Rightarrow h(x).f(x) \geq \varepsilon f(x) \quad (\because f(x) > 0)$$
$$\Rightarrow \int_S h(x).f(x)\,dx \geq \varepsilon \int_S f(x)\,dx \tag{1}$$

But by definition, we have

$$E(h(X)) = \int_{-\infty}^{\infty} h(x) f(x)\,dx \quad = \int_S h(x) f(x)\,dx + \int_{R-S} h(x) f(x)\,dx \tag{2}$$

$\sin ce \quad h(x) > 0$ and $f(x) > 0$,

we have $\displaystyle\int_{R-S} h(x).f(x)\,dx \geq 0$ (3)

From (2) and (3), we have

$$E(h(x)) \geq \int_S h(x).f(x)\,dx \quad \geq \varepsilon \int_S f(x)\,dx \qquad \text{(using (1))} \tag{4}$$

But we have $S = \{x \in R / h(x) \geq \varepsilon\}$

Therefore,

$$\int_S f(x)\,dx = P\{x \in R / h(x) \geq \varepsilon\} \quad = P\{h(x) \geq \varepsilon\} \tag{5}$$

Thus, from (4) and (5), we may write

$$E\{h(x)\} \geq \varepsilon P\{h(x) \geq \varepsilon\} \;\Rightarrow P\{h(x) \geq \varepsilon\} \leq \frac{E\{h(x)\}}{\varepsilon}$$

Hence, the proof.

Deducing Markov's inequality :

Let $h(x) = |X|^r$ and $\varepsilon = k^r$, where $r > 0$ and k>0 in the above inequality, we get

$$P\{|X|^r \geq k^r\} \leq \frac{E\{|X|^r\}}{k^r} \qquad \text{i.e} \qquad P\{|X| \geq k\} \leq \frac{E\{|X|^r\}}{k^r}$$

6.7 SCHWARTZ INEQUALITY:

Putting $r = s = 2$ in holder's inequality we get

$$\mathrm{E}(|\mathrm{XY}|) \leq \sqrt{\mathrm{E}(|\mathrm{X}|^2)\mathrm{E}(|\mathrm{Y}|^2)} \qquad \text{(or)} \quad \mathrm{E}(|\mathrm{XY}|) \leq \sqrt{\mathrm{E}(X^2)\mathrm{E}(Y^2)} \tag{1}$$

Note: For any two random variables we have

$$\left|\mathrm{E}(\mathrm{XY})\right| \leq \mathrm{E}(|\mathrm{XY}|) \tag{2}$$

From (1) and (2) we have

$$\left|\mathrm{E}(\mathrm{XY})\right| \leq \sqrt{\mathrm{E}\left(X^2\right)\mathrm{E}\left(Y^2\right)}$$

Replacing X by $X-E(X)$ and Y by $Y-E(Y)$ in the above Schwartz inequality, we get

$$\left|E\left[X-E(X)\left(Y-E(Y)\right)\right]\right| \leq \sqrt{E\left[X-E(X)\right]^2 . E\left[Y-E(Y)\right]^2}$$

$$\text{i.e.,} \quad \left|\operatorname{cov}(X,Y)\right| \leq \sqrt{V(X)V(Y)} \tag{3}$$

The correction coefficient between X and Y is defined as

$$\mathrm{r_{XY}} = \frac{\left|\operatorname{cov}(X,Y)\right|}{\sqrt{V(X)V(Y)}}$$

From equation (3), we may observe this

$$|\mathrm{r_{XY}}| \leq 1$$

6.8 KOLMOGOROV'S INEQUALITY:

Statement: Let $X_1,X_2,...,X_n$ be independent RVs with common mean 0 and variance σ_k^2, k=1,2,…,n, respectively. Then for any ε >0,

$$P\left\{\max_{1\leq k\leq n} |S_k| > \varepsilon\right\} \leq \sum_1^n \sigma_i^2 / \varepsilon^2, \quad \text{where } S_k = \sum_{i=1}^k X_i$$

Proof: Let

$$A_k = \left\{\max_{1\leq j\leq k} |S_j| \leq \varepsilon\right\} = \left\{\max\left[|S_1|,|S_2|,...,|S_k|\right] \leq \varepsilon\right\}$$
$$= \left\{|S_1| \leq \varepsilon, |S_2| \leq \varepsilon, ..., |S_k| \leq \varepsilon\right\} \tag{1}$$
$$= \left\{|S_1| \leq \varepsilon \cap |S_2| \leq \varepsilon \cap ... \cap |S_k| \leq \varepsilon\right\}, \quad k = 1,2,...,n$$

$$\Rightarrow A_{k+1} = \left\{|S_1| \leq \varepsilon \cap |S_2| \leq \varepsilon \cap ... \cap |S_k| \leq \varepsilon \cap |S_{k+1}| \leq \varepsilon\right\}$$
$$\Rightarrow \quad A_{k+1} = A_k \cap \left\{\cap |S_{k+1}| \leq \varepsilon\right\}$$
$$\Rightarrow \mathrm{A}_{k+1} \subset \mathrm{A_k} \quad \Rightarrow \quad \left\{\mathrm{A_k}\right\}\downarrow$$

If $\overline{A}_k$ is the complement of A_k, then it is obviously $\left\{\overline{\mathrm{A}}_\mathrm{k}\right\}\uparrow$ and

$$\overline{A_k} = \{|S_1| > \varepsilon \cup |S_2| > \varepsilon \cup \ldots \cup |S_k| > \varepsilon\} = \left\{\overline{A}_{k-1} \cup |S_k| > \varepsilon\right\} = \{|S_k| > \varepsilon\}$$

$\left(\text{since } \overline{A}_{k-1} \subseteq \overline{A_k}\right)$

Denote $A_0 = \Omega$ and

$$B_k = A_{k-1} \cap \overline{A_k} = \{|S_1| \le \varepsilon, \ldots, |S_{k-1}| \le \varepsilon\} \cap \{|S_k| > \varepsilon\}$$
$$= \{|S_1| \le \varepsilon, \ldots, |S_{k-1}| \le \varepsilon, |S_k| > \varepsilon\}$$

In particular

$$B_1 = A_0 \cap \overline{A_1} = \Omega \cap \{|S_1| > \varepsilon\} = \{|S_1| > \varepsilon\}$$
$$B_2 = A_1 \cap \overline{A_2} = \{|S_1| \le \varepsilon \cap |S_2| > \varepsilon\}$$
$$B_3 = A_2 \cap \overline{A_3} = \{|S_1| \le \varepsilon \cap |S_2| \le \varepsilon \cap |S_3| > \varepsilon\}$$

So that B_k 's are disjoint, and it follows that

$$B_1 + B_2 = \{|S_1| > \varepsilon\} + \{|S_1| \le \varepsilon \cap |S_2| > \varepsilon\}$$
$$= \{|S_1| > \varepsilon + |S_1| \le \varepsilon\} \cap \{|S_1| > \varepsilon + |S_2| > \varepsilon\}$$
$$= \Omega \cap \{|S_1| > \varepsilon \cup |S_2| > \varepsilon\} = \{|S_1| > \varepsilon \cup |S_2| > \varepsilon\} = \{|S_2| > \varepsilon\}$$

Now, we have

$$\overline{A}_n = \{|S_n| > \varepsilon\} = \sum_{k=1}^{n} B_k \quad \text{and} \quad B_k = A_{k-1} \cap \overline{A_k}$$

$$\text{since } A_{k-1} = \{|S_1| \le \varepsilon, |S_2| \le \varepsilon, \ldots, |S_{k-1}| \le \varepsilon\} \subseteq \{|S_{k-1}| \le \varepsilon\} \tag{2}$$

$$B_k \subseteq \{|S_{k-1}| \le \varepsilon\} \cap \overline{A_k} = \{|S_{k-1}| \le \varepsilon \cap |S_k| > \varepsilon\}$$

Let us denote I_{B_k} as an indicator function of the event B_k defined as

$$I_{B_k}(\omega) = 1 \text{ if } \omega \in B_k$$
$$= 0 \text{ if } \omega \notin B_k$$

Then

$$E(S_n I_{B_k})^2 = E\left\{(S_n - S_k) I_{B_k} + S_k I_{B_k}\right\}^2$$
$$= E\left\{(S_n - S_k)^2 I_{B_k} + S_k^2 I_{B_k} + 2S_k (S_n - S_k) I_{B_k}\right\} \quad (\text{since } I_{B_k}^2 = I_{B_k})$$

Since $S_n - S_k = X_{k+1} + \ldots + X_n$ and $S_k I_{B_k}$ are independent,

since S_k depending on the set of variables $x_1, x_2, x_3, \ldots, x_k$ and S_n-S_k is depending on another set of variables $x_{k+1}, \ldots, x_n$, the varibales S_k and S_n-S_k are mutually exclusive

$\therefore$ The variables s_n-s_k and s_k are independent

$$\text{cov}((S_n\text{-}S_k), S_k) = 0 \Rightarrow E\left[(S_n\text{-}S_k) S_k\right] - E(S_n - S_k) E(S_k) = 0 \Rightarrow E\left[(S_n\text{-}S_k) S_k\right] = 0$$

(since $E\left(X_k\right)=0$ and hence $\mathrm{E}\left[S_k\right]=0$ for all k,)
Thus, we have

$$E(S_nI_{B_k})^2 = E\left\{\left(S_n - S_k\right)I_{B_k}\right\}^2 + E\left\{S_kI_{B_k}\right\}^2 \geq E\left\{S_kI_{B_k}\right\}^2 = E\left\{S_k^2 I_{B_k}\right\}$$

$$= \int_{\mathrm{R}} S_k^2 I_{B_k} = \int_{\mathrm{B_k}} S_k^2 I_{B_k} + \int_{\overline{\mathrm{B_k}}} S_k^2 I_{B_k} = \int_{\mathrm{B_k}} \mathrm{S}_k{}^2 + 0 = \int_{\mathrm{B_k}} \mathrm{S}_k{}^2 = \varepsilon^2 P(B_k) \tag{3}$$

$$\Rightarrow \sum_{k=1}^{n} E(S_nI_{B_k})^2 = E\left(S_n^2 I_{\overline{A}_n}\right) \leq E(S_n^2) = \sum_{1}^{n} \sigma_k^2$$

$$\Rightarrow \sum_{1}^{n} \sigma_k^2 \geq \sum_{k=1}^{n} E(S_nI_{B_k})^2 \tag{4}$$

From Eqs. (3) & (4), we have

$$\sum_{1}^{n} \sigma_k^2 \geq \varepsilon^2 \sum_{1}^{n} P\left(\mathrm{B}_k\right) = \varepsilon^2 P\left(\sum_{1}^{\mathrm{n}} \mathrm{B}_k\right) = \varepsilon^2 P(\overline{\mathrm{A}}_n) \qquad \text{(from eq(2))} \tag{5}$$

From Eq.(1), we have $A_n = \left\{\max\limits_{1\leq j\leq n} |S_j| \leq \varepsilon\right\}$

$$\therefore \overline{A}_n = \left\{\max_{1\leq j\leq n} |S_j| > \varepsilon\right\} = \left\{\max_{1\leq k\leq n} |S_k| > \varepsilon\right\} \tag{6}$$

Substituting (6) in (5) we get

$$\sum_{1}^{n} \sigma_k^2 \geq \varepsilon^2 P\left\{\max_{1\leq k\leq n} |S_k| > \varepsilon\right\} \Rightarrow P\left\{\max_{1\leq k\leq n} |S_k| > \varepsilon\right\} \leq \sum_{1}^{n} \sigma_k^2 \Big/ \varepsilon^2$$

Hence the proof.

Note: Take n=1; then $P\left\{|X_1| > \varepsilon\right\} \leq \dfrac{\sigma_1^2}{\varepsilon^2}$, which is Chebychev's inequality.

6.9 HAJEK-RENYI INEQUALITY:

Statement: Let $\mathrm{X}_1, \mathrm{X}_2, \ldots$ be independent random variables such that $E\left(X_i\right)=0$ and $V\left(X_i\right)=\sigma_i<\infty$. If $C_1, C_2, \ldots$ i non-increasing sequence of positive constants, then for any positive integers m, n with $m<n$ and arbitrary $\in>0$,

$$P\left(\max C_k\left|X_1+X_2+\ldots+X_k\right| \geq \in\right) \leq \frac{1}{\in^2}\left(C_m^2 \sum_{i}^{m} \sigma_i^2 + \sum_{m+1}^{n} C_i^2 \sigma_i^2\right)$$

Proof: To prove the inequality, consider the Quantity

$$Y = \sum_{k=m}^{n-1} S_k^2\left(C_k^2 - C_{k+1}^2\right) + C_n^2 S_n^2$$

Where $S_k = X_1 + X_2 + \ldots + X_k$. It is easy to show that

$$E(Y) = C_m^2 \sum_{k=1}^{m} \sigma_k^2 + \sum_{k=m+1}^{n} C_k^2 \sigma_k^2$$

Let $E_i\left(i = m, m+1, ..., n\right)$ be the event $C_j\left|S_j\right| < \in$ for $m \le j < i$ and $C_i\left|S_i\right| \ge \in$. the events E_i , are mutually exclusive and $P\left\{\max_{m \le k < n} C_k \left|S_k\right| \ge \in\right\} = \sum_{k=m}^{n} P\left(E_i\right)$

Let E_o denote that event that $C_j\left|S_j\right| < \in$ for $m \le j \le n$. Then by the definition of condition expectation $E(Y) = \sum_{i=0}^{n} E\left(Y \setminus E_i\right) P\left\{E_i\right\} \ge \sum_{i=1}^{n} E\left(Y \setminus E_i\right) P\left\{E_i\right\}$ for $k \ge i,$

$$E\left(S_k^2 \mid E_i\right) = E\left[\left\{S_i^2 + \left(X_{i+1} + ... + X_k\right)^2 + 2S_i\left(X_{i+1} + ... + X_k\right)\right\} \mid E_i\right]$$

$$\ge E\left(S_k^2 \mid E_i\right) + 2E\left(S_i\left(X_{i+1} + ... + X_k\right) \mid E_i\right)$$

But the occurrence of the event E_i only imposes a retraction on the first i if the variables X_i and the following variables, under this condition, remain independent of one another and of S_i. Hence, for $j \ge i, E\left(S_i X_j \mid E_i\right) = 0,$ thus giving the inequality $E\left(S_k^2 \mid E_i\right) \ge E\left(S_i^2 \mid E_i\right).$ When E_i is given $\left|S_i\right| \ge \in \mid C_i$ .But

$$E\left(Y \setminus E_i\right) = \sum_{k=m}^{n-1} E\left(S_k^2 \mid E_i\right)\left(C_k^2 - C_{k+1}^2\right) + C_n^2 E\left(S_k^2 \mid E_i\right)$$

$$\ge \sum_{k=i}^{n-1} E\left(S_k^2 \mid E_i\right)\left(C_k^2 - C_{k+1}^2\right) + C_n^2 E\left(S_k^2 \mid E_i\right)$$

$$\ge \frac{\in^2}{C_i^2}\left[\sum_{k=i}^{n-1}\left(C_k^2 - C_{k+1}^2\right) + C_n^2\right] = \in^2$$

And there $E(Y) \ge \in^2 \sum_{k=m}^{n} P\left(E_i\right).$ the required result follows from the exact expression derived for $E(Y)$.

Kolmogorov's inequality of example 3.2 follows, as a special case choosing $m = 1, C_1 = C_2 = ... = C_n = 1.$

6.10 CONCLUSION:

The study of **probability inequalities and the inversion theorem** plays a crucial role in probability theory, statistics, and mathematical analysis. These concepts provide essential tools for bounding probabilities, analyzing convergence, and solving optimization problems.

a) **Inversion Theorem**: This theorem allows us to uniquely recover a probability distribution from its characteristic function. It is particularly useful in proving limit theorems and studying the behavior of probability distributions.
b) **Chebyshev's Inequality**: A fundamental result that provides an upper bound on the probability of a random variable deviating significantly from its mean. It is widely used in statistics and probability theory for variance-based estimates.
c) **Jensen's Inequality**: A powerful inequality that relates convex functions to expectations, playing a key role in optimization, information theory, and machine learning.
d) **Cauchy-Schwarz Inequality**: A fundamental result in linear algebra and probability, ensuring that the correlation between two random variables does not exceed the product of their standard deviations.
e) **Minkowski's Inequality**: An extension of the triangle inequality to spaces of random variables, useful in functional analysis and statistical applications.
f) **Markov's Inequality**: Provides a general bound on the probability that a non-negative random variable exceeds a given value. It is used in proving other inequalities such as Chebyshev's.
g) **Schwarz Inequality**: A special case of the Cauchy-Schwarz inequality, applied in probability theory and linear algebra.
h) **Kolmogorov's Inequality**: Provides bounds on the probability of maximum deviations of sums of independent random variables, crucial in stochastic processes.
i) **Hajek-Rényi Inequality**: A result in probability theory that provides bounds on the probability of deviations for sequences of independent random variables.

6.11 SELF ASSESSMENT QUESTIONS:

1) Prove Chebyshev's inequality and explain how it is used in probability theory.
2) Show that Jensen's inequality holds for a convex function and a random variable.
3) Prove the Cauchy-Schwarz inequality for random variables.
4) Derive Minkowski's inequality and explain its importance in probability and statistics.
5) Explain how Kolmogorov's inequality can be used in probability limit theorems.
6) Discuss the applications of Hajek-Rényi inequality in probability theory.
7) Explain the inversion theorem and prove how a characteristic function uniquely determines a probability distribution.

6.12 SUGGESTED READINGS

1. Modern probability theory by B. R. Bhat, Wiley Eastern Limited.
2. An introduction to probability theory and mathematical statistics by V. K. Rohatgi, John Wiley.
3. AnOutlineofstatisticstheory-1, by A.M.GOON, M.K. Gupta and B. Das gupta, the World Press Private Limited, Calcutta.
4. The Theory of Probability by B.V. Gnedenko, MIR Publishers, Moscow.
5. Discrete distributions -N.L. Johnson and S. Kotz, John wiley & Sons.
6. ContinuousUnivariatedistributions,vol.1&2N.L.JohnsonandS.Kotz,JohnWiley&Sons.
7. Mathematical Statistics-Parimal Mukopadhyay, New Central Book Agency (P) Ltd., Calcutta.

Dr. Syed Jilani

LESSON -7

CONVERGENCE OF SEQUENCE OF RANDOM VARIABLES

OBJECTIVES :

After studying this unit, you should be able to:

- To understanding the convergence of sequence of random variables.
- To know the concept of Structure and convergence of sequence of random variables.
- To acquire knowledge about significance of convergence of sequence of random variables.
- To understand the purpose and objectives of pivotal provisions of the convergence of sequence of random variables.

STRUCTURE:

7.1. INTRODUCTION:

In probability theory, the concept of **convergence of random variables** is crucial in understanding how sequences of random variables behave as they approach a limiting random variable. Different **modes of convergence** provide various ways to measure this behavior, each with different levels of strictness.

7.2 CONVERGENCE IN PROBABILITY

Let $\{X_n\}$ be a sequence of Random variables defined on some probability space (Ω, S, P) we say that the sequence $\{X_n\}$ converges in probability to the random variable 'X' if for every $\varepsilon > 0$,

$P\{\omega \in \Omega / |X_n(\omega) - X(\omega)| > \varepsilon\} \to 0$ as $n \to \infty$ or simply $\lim_{n\to\infty} P\{\omega \in \Omega / |X_n(\omega) - X(\omega)| > \varepsilon\} = 0$

Equivalently, if for every ε > 0,

$$P\{|X_n - X| < \varepsilon\} \to 1 \quad \text{as } n \to \infty \quad \text{i.e., } \lim_{n\to\infty} P\{|X_n - X| < \varepsilon\} = 1$$

More elaborately, a sequence of RVs $\{X_n\}$ defined on some probability space (Ω, S, P) is said to converge in probability to a RV X if for every given $\varepsilon > 0 \;\&\; \delta > 0, \exists$ a N such that

$$P\{|X_n - X| > \varepsilon\} < \delta \;\; \forall \; n \geq N$$

We denote convergence of $\{X_n\}$ in probability to X as $X_n \xrightarrow{P} X$. OR Plim $X_n = X$

We may also say that $\{X_n\}$ is said to convergence stochastically to X.

Remark : This concept plays an important role in statistics. Consistency of estimators and weak law of large numbers are instances of this concept. Intuitively, it means that the difference between X_n & X is likely to be small with large probability for large 'n'.

7.3 .CONVERGENCE ALMOST SURELY:

Let $\{X_n\}$ be a sequence of random variables defined on some probability space (Ω, S, P) we say that the sequence $\{X_n\}$ converges almost surely to a random variable 'X' if and only if,

$$P\left\{\omega \in \Omega / \lim_{n\to\infty} |X_n(\omega) - X(\omega)| < \varepsilon\right\} = 1$$

i.e. $P\left\{\omega \in \Omega / \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\} = 1$ or simply $P\left\{\lim_{n\to\infty} X_n = X\right\} = 1$

This is denoted as $\lim_{n\to\infty} X_n = X$ almost surely or $X_n \xrightarrow{a.s} X$

<u>Note</u> : In this case, the set of convergence of $\{X_n\}$ has unit probability. Thus, <u>lim</u> X_n & X are equivalent random variables.

7.4. CONVERGENCE IN LAW

Let $\{X_n\}$ be a sequence of random variables defined on some probability space (Ω, S, P) and $\{F_n\}$ is the corresponding sequence of distribution functions. Now we say that the $\{X_n\}$ convergence in law (distribution) to a random variable X (defined on the same probability space (Ω, S, P)) if it exists with a distribution function F such that

$$F_n(x) \to F(x) \quad \text{as} \quad n \to \infty \text{ at every point 'x' at which F is continuous.}$$

We may also say that $\{F_n\}$ convergence in law or weakly to F and we write $F_n \xrightarrow{w} F$ or $X_n \xrightarrow{L} X$

7.5 CONVERGENCE IN THE r^{th} MEAN:

A sequence of random variable $\{X_n\}$ defined on some probability space (Ω, S, P) such that $E(|X_n|^r) < \infty$ for some $r > 0$, is said to be convergence in the r^{th} mean to a RV X if $E\left(|X_n|^r\right) < \infty$ and $E\left(|X_n - X|^r\right) \to 0$ as $n \to \infty$ we denote it as $X_n \xrightarrow{L_r} X$

Note : When r=1, convergence in r^{th} mean is called as convergence in mean.

7.5.1. Convergence in mean square or quadratic mean

A sequence of random variables $\{X_n\}$ defined on some probability space (Ω, S, P) such that $E(X_n^2) < \infty$ is said to converge in mean square (quadratic mean) to a random variable 'X' if $E(X^2) < \infty$ and $E[(X_n-X)^2] \to 0$ as $n \to \infty$. We denote it as $X_n \xrightarrow{L_2} X$.

Thus convergence in mean square or quadratic mean is a special case of convergence in r^{th} mean when r = 2.

7.6 RELATIONSHIP BETWEEN CONVERGENCE IN PROBABILITY AND CONVERGENCE IN LAW:

Statement: Convergence in probability $\Rightarrow$ convergence in law. i.e., $X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{L} X$.

Proof:

Since, we have given convergence in probability, it immediately follows, for given $\varepsilon > 0$

$$\lim_{n\to\infty} P\left(|X_n - X| > \varepsilon\right) = 0 \tag{1}$$

Let x be a continuous point of F. Then, for any arbitrary $\varepsilon > 0$, we have

$$\{\omega \in \Omega / X_n(\omega) \le x\} = \{X_n \le x \cap X \le x + \varepsilon\} + \{X_n \le x \cap X > x + \varepsilon\}$$

Taking probability on both sides

$$\begin{aligned} F_n(x) = P(X_n \le x) &= P(X_n \le x, X \le x+\varepsilon) + P(X_n \le x, X > x+\varepsilon) \\ &\le P(X \le x+\varepsilon) + P(X_n \le x, X > x+\varepsilon) \end{aligned}$$

But, we may write

$$\begin{aligned} \{X_n \le x, X > x+\varepsilon\} &= \{-X_n \ge -x, X > x+\varepsilon\} \cup \{X_n \le x, -X \le -x-\varepsilon\} \\ &= \{X - X_n > \varepsilon \text{ or } X_n - X \le -\varepsilon\} = \{|X_n - X| > \varepsilon\} \end{aligned}$$

Therefore, we have

$$F_n(x) \le P\left(X \le x+\varepsilon\right) + P\left(\left|X_n - X\right| > \varepsilon\right) = F\left(x+\varepsilon\right) + P\left(\left|X_n - X\right| > \varepsilon\right) \tag{2}$$

Similarly, we can show that

$$F\left(x-\varepsilon\right) = P\left(X \le x-\varepsilon\right) = P\left(X \le x-\varepsilon, X_n \le x\right) + P\left(X \le x-\varepsilon, X_n > x\right)$$

$$\le F_n(x) + P\left(\left|X_n - X\right| > \varepsilon\right) \tag{3}$$

From (2) and (3), we have

$$F\left(x-\varepsilon\right) - P\left(\left|X_n - X\right| > \varepsilon\right) \le F_n(x) \le F\left(x+\varepsilon\right) + P\left(\left|X_n - X\right| > \varepsilon\right)$$

Taking the limit $n \to \infty$ we get

$$F\left(x-\varepsilon\right) \le \lim_{n\to\infty} F_n(x) \le F\left(x+\varepsilon\right) \qquad \text{(from Eq. (1))}$$

This holds for all $\varepsilon>0$. Take the limit as $\varepsilon \to 0$ and use the fact that F is continuous at x, and then we get

$$\lim_{n\to\infty} F_n(x) = F\left(x\right)$$

Hence the proof.

Fix $\varepsilon>0$. Then,

$$P\left(\left|X_n - c\right| > \varepsilon\right) = P\left(X_n < c-\varepsilon\right) + P\left(X_n > c+\varepsilon\right)$$

$$\le P\left(X_n \le c-\varepsilon\right) + P\left(X_n > c+\varepsilon\right)$$

$$= F_n\left(c-\varepsilon\right) + 1 - F_n\left(c+\varepsilon\right)$$

$$\to F\left(c-\varepsilon\right) + 1 - F\left(c+\varepsilon\right)$$

$$= 0 + 1 - 1 = 0$$

A useful Result: $X_n \xrightarrow{a.s.} X \Leftrightarrow \lim\limits_{n0\to\infty} P\left\{\sup_{n\ge n0}\left|X_n - X\right| > \varepsilon\right\} = 0$ for all $\varepsilon > 0$.

Proof: Since $X_n \xrightarrow{a.s.} X$, $X_n - X \xrightarrow{a.s.} 0$, and it will be sufficient to show the equivalence

of
a) $X_n \xrightarrow{a.s.} 0$ and
b) $\lim\limits_{n0\to\infty} P\left\{\sup_{n\ge n0}\left|X_n\right| > \varepsilon\right\} = 0.$

Let us suppose that (a) holds. Let $\varepsilon > 0$, and write

$$A_n\left(\varepsilon\right) = \left\{\sup_{m\ge n}\left|X_m\right| > \varepsilon\right\} \text{ and } C = \left\{\lim_{n\to\infty} X_n = 0\right\}.$$

Also write $B_n(\varepsilon) = \mathrm{C} \cap A_n(\varepsilon)$, and note that $B_{n+1}(\varepsilon) \subset B_n(\varepsilon)$, and the limit set $\cap_{n=1}^{\infty} B_n(\varepsilon) = \Phi$. It follows that

$$\lim_{n\to\infty} PB_n(\varepsilon) = P\left\{\cap_{n=1}^{\infty} B_n(\varepsilon)\right\} = 0.$$

Since $PC = 1,\ PC^c = 0,$ and we have

$$\begin{aligned} PB_n(\varepsilon) &= P\{A_n \cap C\} = 1 - P\left\{C^c \cup A_n^c\right\} \\ &= 1 - PC^c - PA_n^c + P\left\{C^c \cap A_n^c\right\} \\ &= PA_n + P\left\{C^c \cap A_n^c\right\} \\ &= PA_n. \end{aligned}$$

It follows that (b) holds.

Conversely, let $\lim\limits_{n\to\infty} PA_n(\varepsilon) = 0,$ and write

$$D(\varepsilon) = \left\{\overline{\lim_{n\to\infty}} |X_n| > \varepsilon > 0\right\}.$$

Since $D(\varepsilon) \subset A_n(\varepsilon)$ for n=1,2,…, it follows that $PD(\varepsilon) = 0.$ Also,

$$\mathrm{C}^c = \left\{\lim_{n\to\infty} X_n \neq 0\right\} \subset \bigcup \left\{\overline{\lim_{n\to\infty}} |X_n| > \frac{1}{k}\right\},$$

So that

$$1 - PC \le \sum_{k=1}^{\infty} PD\left(\frac{1}{k}\right) = 0,$$

Hence the proof.

7.7 RELATIONSHIP BETWEEN ALMOST SURELY CONVERGENT AND CONVERGENT IN PROBABILITY:

Statement: Almost surely convergence $\Rightarrow$ convergence in probability . Symbolically,

$$X_n \xrightarrow{a.s.} X \Rightarrow X_n \xrightarrow{P} X.$$

That is if $\{X_n\}$ be a sequence of RVs defined on some probability space (Ω, S, P) converges almost surely to the random variable 'X' (defined on the same (Ω, S, P)), then $\{X_n\}$ converges to X in probability .

Proof:

We have the result,

$$X_n \xrightarrow{a.s.} X \quad \Rightarrow \quad \lim_{n0 \to \infty} P\left\{\sup_{n \geq n0} \left| X_n - X \right| > \varepsilon \right\} = 0 \text{ for all } \varepsilon > 0.$$

Indeed, we can write equivalently that for given $\varepsilon > 0, \eta > 0$, $\exists\ n_0 = n_0(\varepsilon, \eta)$ such that

$$P\left\{ \sup_{n \geq n_0} \left| X_n - X \right| > \varepsilon \right\} < \eta \Rightarrow P\left\{ \bigcup_{n \geq n_0} \left\{ \left| X_n - X \right| > \varepsilon \right\} \right\} < \eta \qquad \text{(By def. of Supremum)}$$

$$\Rightarrow P\left[\bigcap_{n=n_0}^{\infty} \left\{ \left| X_n - X \right| \leq \varepsilon \right\} \right] \geq 1 - \eta$$

$$\because \left(\bigcup_{n \geq n_0} \left\{ \left| X_n - X \right| > \varepsilon \right\} \right)^C = \bigcap_{n=n_0}^{\infty} \left\{ \left| X_n - X \right| \leq \varepsilon \right\}$$

Clearly,

$$\bigcap_{n=n_0}^{\infty} \left\{ \left| X_n - X \right| \leq \varepsilon \right\} \subset \left\{ \left| X_n - X \right| \leq \varepsilon \right\} \quad \text{for} \quad n \geq n_0.$$

Then, it follows that for $n \geq n_0$, $\quad P\left\{ \left| X_n - X \right| \leq \varepsilon \right\} \geq P\left[\bigcap_{n=n_0}^{\infty} \left\{ \left| X_n - X \right| \leq \varepsilon \right\} \right] \geq 1 - \eta$

$$\Rightarrow P\left\{ \left| X_n - X \right| > \varepsilon \right\} < \eta$$

which is the same as saying that $X_n \xrightarrow{P} X$.

Note: The above results implied Convergence in distribution is weaker than convergence in probability and hence Convergence almost Surely.

7.8 RELATIONSHIP BETWEEN MEAN SQUARE CONVERGENCE AND CONVERGENCE IN PROBABILITY:

Mean square convergence implies convergence in probability from "Morkove's inequality". We know that for any non-negative function h(x) of any positive real numbers ε and then $P\{h(x) > \varepsilon\} \leq \dfrac{E[h(x)]}{\varepsilon}$ taking $h(X) = (X_n - c)^2$ and E as E^2 in equation

$$P\{(X_n - c)^2 > \varepsilon^2\} \leq \frac{E[X_n - c]^2}{\varepsilon^2}$$

Applying positive square root to the probability. We get,

$$P\{(X_n - c) > \varepsilon\} \leq \lim_{n \to \infty} \frac{E[X_n - c]^2}{\varepsilon^2}$$

By hypothesis X_n converges to zero $(X_n \to 0)$ in mean square i.e., $\lim_{n\to\infty} \frac{E[X_n - c]^2}{\varepsilon^2} = 0$

then obviously $P\{(X_n - c) > \varepsilon\} = 0$.

Hence mean square convergence implies convergence in probability.

7.9 CONCLUSION:

The concept of **convergence of random variables** is fundamental in probability theory and statistical inference. Different modes of convergence describe how sequences of random variables behave as they approach a limit, which is crucial in areas like asymptotic analysis, statistical estimation, and stochastic processes.

a) **Convergence in Probability**: A sequence of random variables converges in probability to a limit if the probability of large deviations decreases as the sample size increases. This is widely used in **statistical estimation and the law of large numbers**.

b) **Convergence Almost Surely**: A stronger form of convergence where the sequence converges for almost all outcomes in the probability space. This is critical in stochastic processes and martingales.

c) **Convergence in Law (Distribution)**: This means that the distribution of a sequence of random variables approaches the distribution of a limiting variable. It is commonly used in proving **central limit theorems**.

d) **Convergence in rth Mean**: A sequence converges in rth mean if the expected value of the absolute difference between the random variables and the limit, raised to the power r, tends to zero. It is useful in **mean-based estimations**.
   - **Mean Square (Quadratic Mean) Convergence**: A special case where r=2, meaning that the expected squared difference between the sequence and the limit tends to zero. It is important in **regression analysis and stochastic modeling**.

e) **Relationships Between Different Modes of Convergence**:
   - **Convergence Almost Surely vs. Convergence in Probability**: Almost sure convergence implies convergence in probability, but the converse is not necessarily true.
   - **Mean Square Convergence vs. Convergence in Probability**: Mean square convergence implies convergence in probability, but the reverse is true only under additional conditions.
   - **Convergence in Probability vs. Convergence in Law**: Convergence in probability implies convergence in law, but not vice versa.

7.10 SELF ASSESSMENT QUESTIONS:

1) Prove that almost sure convergence implies convergence in probability.
2) Explain how convergence in probability implies convergence in law, but not vice versa.
3) Show that convergence in mean square implies convergence in probability.
4) Provide an example where convergence in probability does not imply almost sure convergence.
5) Compare and contrast different types of convergence in probability theory.

6) Given a sequence of random variables Xn converging to X in mean square, prove that it also converges in probability.
7) Explain how the different modes of convergence are used in real-world applications, such as statistical inference or machine learning.

7.11 SUGGESTED READING BOOKS:

1. Modern probability theory by B. R. Bhat, Wiley EasternLimited.
2. An introduction to probability theory and mathematical statistics by V. K. Rohatgi, John Wiley.
3. AnOutlineofstatisticstheory-1, by A.M.GOON, M.K. Gupta and B. Das gupta, the World Press Private Limited, Calcutta.
4. The Theory of Probability by B.V. Gnedenko, MIR Publishers, Moscow.
5. Discrete distributions -N.L. Johnson and S. Kotz, John wiley & Sons.
6. ContinuousUnivariatedistributions,vol.1&2N.L.JohnsonandS.Kotz,JohnWiley&Sons.
7. Mathematical Statistics-Parimal Mukopadhyay, New Central Book Agency (P) Ltd., Calcutta.

Dr. Syed Jilani

LESSON -8

WEAK LAW OF LARGE NUMBERS AND STRONG LAW OF LARGE NUMBERS

OBJECTIVES:

After studying this unit, you should be able to:

- To understanding the Weak law of large numbers and Strong law of large numbers
- To know the concept of Structure and Weak law of large numbers and Strong law of large numbers
- To acquire knowledge about significance of Weak law of large numbers and Strong law of large numbers
- To understand the purpose and objectives of pivotal provisions of the Weak law of large numbers and Strong law of large numbers

STRUCTURE

8.1 INTRODUCTION:

The Law of Large Numbers (LLN) is a fundamental theorem in probability theory that explains how the average of a sequence of independent and identically distributed (i.i.d.) random variables behaves as the sample size increases. It states that, under certain conditions, the sample average converges to the expected value of the underlying distribution.

There are two main versions of the LLN:

1. **Weak Law of Large Numbers (WLLN)** – Ensures convergence in probability.
2. **Strong Law of Large Numbers (SLLN)** – Ensures almost sure convergence.

Both versions express the same idea: as the number of observations increases, the sample mean becomes a more accurate estimate of the population mean. However, they differ in the strength of their convergence guarantees.

8.2 WEAK LAW OF LARGE NUMBERS:

Let $\{X_n\}$ be a sequence of RVs. Write $S_n = \sum_{k=1}^{n} X_k, k = 1,2....$ In this section we may answer the following question in the affirmative:

Do there exist sequence of constants A_n and $B_n>0, B_n \to \infty$ as $n \to \infty$, such that the sequence of RVs $B_n^{-1}(S_n - A_n)$ converges in probability to 0 as $n \to \infty$?

Def1: Let $\{X_n\}$ be a sequence of RVs and $S_n = \sum_{k=1}^{n} X_k$, $n=1,2...$ we say that $\{X_n\}$ obeys the weak law of large numbers(WLLN) with respect to the sequence of constants $\{B_n\}, B_n > 0, B_n \uparrow \infty$, if there exists a sequence of real constants A_n such that

$$B_n^{-1}(S_n - A_n) \xrightarrow{P} 0 \quad \text{as } n \to \infty$$

Here, A_n's are called as centering constants and B_n's are as norming constants.

Def2: Let $\{X_n\}$ be a sequence of RVs with $E[X_n] = a_n$. The arithmetic means of $\{X_n\}$ and $\{a_n\}$ are $\overline{X_n} = \frac{1}{n} \sum_{i=1}^{n} X_i$ and $\overline{a_n} = \frac{1}{n} \sum_{i=1}^{n} a_i$

Now, if $\overline{X_n} - \overline{a_n} \xrightarrow{p} 0$, then we say that then $\{X_n\}$ obeys WLLN.

8.3 NECESSARY AND SUFFICIENT CONDITION OF W.L.L.N:

Theorem 1:

Statement: Let $\{X_n\}$ be any sequence of R.V's write $Y_n = n^{-1}\sum_{k=1}^{n} X_k$. A necessary and sufficient condition for the sequence $\{X_n\}$ to satisfy the weak law of large numbers is that

$$E\left\{\frac{Y^2}{1+Y_n^2}\right\} \to 0 \quad \text{as } n \to \infty$$

Proof:

Necessary condition:

Here We have given

$$E\left\{\frac{Y^2}{1+Y_n^2}\right\} \to 0 \quad \text{as } n \to \infty \text{ i.e. } \lim_{n\to\infty} E\left\{\frac{Y^2}{1+Y_n^2}\right\} = 0 \tag{1}$$

For any two positive numbers $a,b,\ a \geq b \geq 0$, we have

$$\frac{a}{1+a}\ \frac{1+\mathrm{b}}{\mathrm{b}} \geq 1 \tag{2}$$

Let $\mathrm{A}=\{\omega \in \Omega / |Y(\omega)| \geq \varepsilon\}$ then $\omega \in \mathrm{A} \Rightarrow |Y_n| \geq \varepsilon \Rightarrow Y_n^{\,2} \geq \varepsilon^2 > 0$

Taking $a = Y_n^{\,2}$ and $b=\varepsilon^2$ in Eq. (2), then we get

$$\frac{Y^2}{1+Y_n^2}\ \frac{1+\varepsilon^2}{\varepsilon^2} \geq 1 \Rightarrow \frac{Y^2}{1+Y_n^2} \geq \frac{\varepsilon^2}{1+\varepsilon^2}$$

Thus, we may write

$$\mathrm{A}=\{\omega \in \Omega / |Y(\omega)| \geq \varepsilon\} = \{\omega \in \Omega / Y_n^{\,2} \geq \varepsilon^2\} = \left\{\omega \in \Omega \Big/ \frac{Y^2}{1+Y_n^2} \geq \frac{\varepsilon^2}{1+\varepsilon^2}\right\}$$

Taking probability on both sides we get $\mathrm{P}(\mathrm{A}) = \mathrm{P}\left\{\frac{Y^2}{1+Y_n^2} \geq \frac{\varepsilon^2}{1+\varepsilon^2}\right\}$

But, we have from Markov's inequality,

$$\mathrm{P}\left\{\frac{Y^2}{1+Y_n^2} \geq \frac{\varepsilon^2}{1+\varepsilon^2}\right\} \leq E\left(\frac{Y^2}{1+Y_n^2}\right) \Big/ \left(\frac{\varepsilon^2}{1+\varepsilon^2}\right) \tag{3}$$

Now, from Eq. (1), it immediately follows

$$\lim_{n\to\infty} \mathrm{P}\left\{\frac{Y^2}{1+Y_n^2} \geq \frac{\varepsilon^2}{1+\varepsilon^2}\right\} = 0 \ \Rightarrow \frac{Y^2}{1+Y_n^2} \xrightarrow{\mathrm{P}} 0 \text{ as } \mathrm{n} \to \infty \Rightarrow Y_n \xrightarrow{\mathrm{P}} 0 \text{ as } \mathrm{n} \to \infty \tag{4}$$

Hence, the necessary condition

Sufficient condition:

For proving the sufficient condition, we have given the sequence $\{X_n\}$ is satisfying the weak law of large numbers. That is we have given

$$Y_n \xrightarrow{\mathrm{P}} 0 \text{ as } \mathrm{n} \to \infty \Rightarrow \textit{for} \text{ a given } \varepsilon>0,\ \lim_{n\to\infty} \mathrm{P}\{|Y_n| > \varepsilon\} = 0 \tag{5}$$

We will prove (4) for the case in which Y_n is of the continuous type. If Y_n has p.d.f. $f_n(\mathrm{y})$, then

$$E\left(\frac{Y^2}{1+Y_n^2}\right)=\int_{-\infty}^{\infty}\frac{y^2}{1+y^2}f_n(\mathrm{y})\,\mathrm{dy}$$

$$=\int_{|\mathrm{y}|>\varepsilon}\frac{y^2}{1+y^2}f_n(\mathrm{y})\,\mathrm{dy}+\int_{|\mathrm{y}|\le\varepsilon}\frac{y^2}{1+y^2}f_n(\mathrm{y})\,\mathrm{dy}$$

$$\le\int_{|\mathrm{y}|>\varepsilon}f_n(\mathrm{y})\,\mathrm{dy}+\int_{|\mathrm{y}|\le\varepsilon}\frac{y^2}{1+y^2}f_n(\mathrm{y})\,\mathrm{dy}\qquad\left(\text{since }\frac{y^2}{1+y^2}\le1\right)$$

$$=P\{|Y_n|>\varepsilon\}+\int_{|\mathrm{y}|\le\varepsilon}\frac{y^2}{1+y^2}f_n(\mathrm{y})\,\mathrm{dy}$$

$$\le\mathrm{P}\{|Y_n|>\varepsilon\}+\frac{\varepsilon^2}{1+\varepsilon^2}\int_{|\mathrm{y}|\le\varepsilon}f_n(\mathrm{y})\,\mathrm{dy}\qquad\left(\text{since }\frac{y^2}{1+y^2}\le\frac{\varepsilon^2}{1+\varepsilon^2}\right)$$

$$\le\mathrm{P}\{|Y_n|>\varepsilon\}\qquad\left(\text{since}\frac{\varepsilon^2}{1+\varepsilon^2}\le1\ \text{ and }\int_{|\mathrm{y}|\le\varepsilon}f_n(\mathrm{y})\,\mathrm{dy}<1\right)$$

Taking $\lim\limits_{n\to\infty}$ on both sides and using Eq. (5),

we get $\lim\limits_{n\to\infty}E\left(\frac{Y^2}{1+Y_n^2}\right)=0$

Hence the sufficient condition

Remark: Since the above theorem (theorem1) applies not to the individual variables but to their sum, it is of limited use. However, all weak laws of large numbers obtained as corollaries of the following theorem.

Theorem 2:

Statement: let $\{X_n\}$ be a sequence of pair wise uncorrelated RVs with $E(X_i)=\mu_i$ and $\mathrm{var}(X_i)=\sigma_i^2,\ i=1,2\ldots.$ If $\sum_{\mathrm{i}=1}^{\mathrm{n}}\sigma_i^2\to\infty\ as\ n\to\infty$, we can choose $\mathrm{A}_n=\sum_{k=1}^{n}\mu_k$ and $B_n=\sum_{i=1}^{n}\sigma_i^2$, that is,

$$\sum_{\mathrm{i}=1}^{\mathrm{n}}\frac{\mathrm{X_i}-\mu_i}{\sum_{i=1}^{n}\sigma_i^2}\xrightarrow{p}0\quad as\quad n\to\infty$$

Proof: We have, by Chebychev's inequality,

$$P\left\{|S_n-\sum_{k=1}^{n}\mu_k|>\varepsilon\sum_{i=1}^{n}\sigma_i^2\right\}\le\frac{E\left[\sum_{i=1}^{n}(X_i-\mu_i)\right]^2}{\varepsilon^2\left(\sum_{i=1}^{n}\sigma_i^2\right)}=\frac{1}{\varepsilon^2\sum_{i=1}^{n}\sigma_i^2}\to 0 \text{ as } n\to\infty.$$

Note1: if the $X_n's$ are identically distributed and pairwise uncorrelated with $E(X_i)=\mu_i$ and $\operatorname{var}(X_i)=\sigma^2<\infty$ we can choose $A_n=n\mu$ and $B_n=n\sigma^2$.

Note2: in the above theorem we can choose $B_n=n$ provided that $n^{-2}\sum_{i=1}^{n}\sigma_i^2\to 0$ as $n\to\infty$.

Note3: in note 1 we can take $A_n=n\mu$ and $B_n=n$, since $\frac{n\sigma^2}{n^2}\to 0$ as $n\to\infty.$ thus if $\{X_n\}$ are pairwise-uncorrelated identically distributed RVs with finite variance, $S_n/n\xrightarrow{p}\mu.$

Example:

Let $(X_1,X_2,\ldots X_n)$ be jointly normal with $E[X_i]=0, E[X_i^2]=1$ for all I and $\operatorname{cov}(X_i,X_j)=\rho$ if $|j-i|=1$, and =0 otherwise, then $S_n=\sum_{k=1}^{n}X_k$ is $N(0,\sigma^2)$ where

$$\sigma^2=\operatorname{var}(S_n)=n+2(n-1)\rho$$

$$E\left\{\frac{Y^2}{1+Y_n^2}\right\}=E\left\{\frac{S_n^{\,2}}{n^2+S_n^{\,2}}\right\}=\frac{2}{\sqrt{2\pi}}\int_0^\infty\frac{x^2}{n^2+x^2}\,e^{\frac{-x^2}{2\sigma^2}}dx$$

$$=\frac{2}{\sqrt{2\pi}}\int_0^\infty\frac{y^2[n+2(n-1)\rho]}{n^2+y^2[n+2(n-1)\rho]}\,e^{\frac{-y^2}{2}}dy$$

$$\le\frac{n+2(n-1)\rho}{n^2}\int_0^\infty\frac{2}{\sqrt{2\pi}}y^2e^{\frac{-y^2}{2}}dy\to 0 \quad \text{as} \quad n\to\infty$$

It follows from Theorem 1 that $n^{-1}S_n\xrightarrow{P}0.$

8.4 CHEBYCHEV'S WLLN:

Theorem-3: Let $\{X_n\}$ be a sequence of pair-wise independent RVs with the following additional hypothesis

(i) $E(X_n)=\mu_n<\infty$

(ii) $V(X_n)=\sigma_n^2\le c\ \forall n=1,2\ldots,$ where $'c'$ is a positive real number.

Then $\{X_n\}$ obeys WLLN centered on μ_n.

Proof: Denote $\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $\overline{\mu_n} = \frac{1}{n}\sum_{i=1}^{n}\mu_i$

Now, by applying Chebychev's inequality to the RV $\overline{X_n}$, then we get

$$P\left\{|\bar{X}_n - E\left(\bar{X}_n\right)| > \varepsilon\right\} < \frac{V(\bar{X}_n)}{\varepsilon^2}$$

$$\Rightarrow P\left\{|\bar{X}_n - \bar{\mu}_n| > \varepsilon\right\} < \frac{1}{\varepsilon^2}\left[V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)\right] = \frac{1}{n^2\varepsilon^2}\sum_{i=1}^{n} V\left(X_i\right) \quad \text{(since } X_i\text{'s are pair-wise independent)}$$

$$= \frac{1}{n^2\varepsilon^2}\sum_{i=1}^{n}\sigma_i^2 < \frac{nc}{n^2\varepsilon^2} = \frac{c}{n\varepsilon^2} \qquad \text{(since } \sigma_i^2 < c\text{)}$$

$$\Rightarrow P\left\{|\bar{X}_n - \bar{\mu}_n| > \varepsilon\right\} < \frac{c}{n\varepsilon^2}$$

Taking limit on both sides, we get

$$\lim_{n\to\infty} P\left\{|\bar{X}_n - \bar{\mu}_n| > \varepsilon\right\} \le \lim_{n\to\infty}\frac{c}{n\varepsilon^2} \Rightarrow \lim_{n\to\infty} P\left\{|\bar{X}_n - \bar{\mu}_n| > \varepsilon\right\} \le \frac{c}{\varepsilon^2}\lim_{n\to\infty}\frac{1}{n}$$

$$\Rightarrow \lim_{n\to\infty} P\left\{|\bar{X}_n - \bar{\mu}_n| > \varepsilon\right\} = 0$$

$$\Rightarrow \bar{X}_n - \bar{\mu}_n \xrightarrow{p} 0$$

Thus $\{X_n\}$ obeys WLLN's centered on μ_n.

Hence, the proof.

8.5 KHINCHIN'S WEAK LAW OF LARGE NUMBERS:

Statement:-Let $x_1, x_2, \ldots$ be a sequence of i.i.d variables. Then $E(X_i) = \mu$ exists and is finite $\Rightarrow \bar{X}_n \xrightarrow{p} \mu$

Proof:- μ be the common finite expectation of all X_n's

$\Rightarrow E(X_n) = \mu \ \forall n.$

All X_n's have the common characteristic function $\phi_X(t)$

$$\Rightarrow \phi_X(t) = 1 + itE(X) + \frac{it^2}{2!}E(X^2) + \cdots + \frac{it^K}{k!}E(X^k) + \cdots$$

$E(X^k)$ is the common k^{th} raw moment of all.

By definition, characteristic function of $\bar{x}_n$

= characteristic function $\left(\frac{x_1}{n}+\frac{x_2}{n}+\cdots+\frac{x_n}{n}\right)$

=Product of characteristic function $\left(\frac{x_1}{n}\cdot\frac{x_2}{n}\cdots\frac{x_n}{n}\right)$

$=$n $^{\text{th}}$ power of characteristic function $(\frac{X}{n})$

(since where $\frac{X}{n}$ is the common random variable represent all of $\frac{x_1}{n}\cdot\frac{x_2}{n}\cdots\frac{x_n}{n}$)

$\Rightarrow$ characteristic function of $\bar{X}_n = n^{th}$ power of C.F 'X' replaced by 't'

$\Rightarrow$ characteristic function of $\bar{X}_n = \left[\phi_X(t/\text{n})\right]^n$

$\Rightarrow$ log of $\bar{X}_n = n\log\phi_X(t/\text{n})$

$$= \text{n } \log(1+\frac{it}{n}E(X)+\frac{it^2}{2!n^2}E(X^2)+\cdots+\frac{it^n}{n!n^n}E(x^n)+\cdots)$$

$$= n(\frac{it}{n}E(X)+\frac{i}{2!}(\frac{t}{n})^2E(X^2)+\cdots)-\frac{1}{2}(\frac{it}{n}E(X)+\cdots+\cdots)^2+\frac{1}{3}(\frac{it}{n}E(X)\frac{it^2}{2!n^2}E(X^2)+\cdots)^3\ldots\ldots)$$

$= (it\mu + \ldots\ldots$ terms in powers of $\text{n}\geq 1$ in denominator)

$\Rightarrow [\lim_{n\to\infty}$ C.F of $\bar{X}_n)] = e^{it\mu}$

Since C.F of $\bar{X}_n = e^{it\mu}$

By Inversion theorem on characteristic function

$\Rightarrow$ DF of $\bar{X}_n \to$ DF of the RV whose CF is $\text{e}^{it\mu}$ as $n\to\infty$

$\Rightarrow$ DF of $\bar{X}_n \to$ DF of one point RV at μ

i.e., $\bar{X}_n \xrightarrow{d} a$ one point random variable at μ

$\Rightarrow \bar{X}_n \xrightarrow{p} \mu$ in probability. Or $(\bar{X}_n-\mu)^p \to 0$

Hence the proof.

8.6 STRONG LAW OF LARGE NUMBERS:

In this section we obtain a stronger form of the law of large numbers discussed in section 6.3. Let $X_1, X_2, ...$ be a sequence of RVs defined on a probability space (Ω, S, P).

Def1: We say that the sequence $\{X_n\}$ obeys the strong law of large numbers(SLLN) with respect to the norming constants $\{B_n\}$ if there exists a sequence of(centering) constants $\{A_n\}$ such that

$$B_n^{-1}(S_n - A_n) \xrightarrow{a.s.} 0 \text{ as } n \to \infty. \qquad (1)$$

We will obtain sufficient conditions for a sequence $\{X_n\}$ to obey the SLLN. In what follows we will be interested mainly in the case $B_n = n.$ Indeed, when we speak of the SLLN we will assume that we are speaking of the norming constants $B_n = n,$ unless specified otherwise.

Def 2:

Let $\{X_n\}$ be a sequence of RVs and $E[X_n] = a_n$. the sequence of arithmetic means of RVs are,

$$\overline{X_n} = \frac{1}{n} \sum_{i=1}^{n} X_i \quad , \quad \overline{a_n} = \frac{1}{n} \sum_{i=1}^{n} a_i$$

If $\left(\overline{X_n} - \overline{a_n}\right) \xrightarrow{as} 0$ then $\{X_n\}$ obeys SLLN.

8.7 KOLMOGOROV SLLN FOR I.I.D CASE:

Statement: If X_k's are independent and identically distributed RVs then $S_n / n \to c\,(a.s.)$, where c is a finite number, iff $E|X| < \infty.$ then $c = EX$.

Proof:

Define $A_n = [|X| \geq n], (n = 0, 1, 2, ...), A_0 = \Omega.$

On $A_n - A_{n+1} = B_n$, $n+1 > |X| \geq n$, so that

$$\sum_0^{\infty} B_n = A_0 = \Omega,$$

$$nPB_n \leq \int_{B_n} |X| < (n+1)PB_n.$$

Summing over n,

$$\sum_1^\infty nPB_n = \sum_1^\infty PA_n \le E|X| \le \sum_0^\infty PA_n = 1 + \sum_1^\infty PA_n. \qquad (1)$$

Suppose $(S_n/n) \to c < \infty (a.s.)$. Then

$$\frac{X_n}{n} = \frac{S_n}{n} - \left(\frac{n-1}{n}\right)\frac{S_{n-1}}{n-1} \to 0(a.s.).$$

By $\sum PA_n < \infty$. from(1), this implies that $E|X| < \infty$.

Conversely suppose $E|X| < \infty$. then from(1) $\sum PA_n < \infty$.

Suppose,

$$X_k^k = X_k, |X_k| < k,$$
$$=0, |X_k| \ge k,$$

i.e., X_k^k is X_k truncated at k and $\overline{S_n} = \sum_1^n X_k^k$. then, since X_k's are identically distributed,

$$\sum_1^\infty \mathrm{P}\left[\mathrm{X_k} \ne X_k^k\right] = \sum_1^\infty P\left[|X_k| \ge k\right] = \sum PA_k < \infty.$$

Hence by lemma, (S_n/n) and $(\overline{S_n}/n)$ are convergence equivalent and it is sufficient if we prove that $(\overline{S_n}/n) \to EX(a.s.)$.

Now since $X_n^n \to X$ and $|X_n^n| \le |X|$ integrable, by dominated convergence theorem, as $n \to \infty$

$$EX_n^n \to EX.$$

By $E(\overline{S_n}/n) = \left(\sum_1^n EX_k^k\right)/n \to EX$, as $n \to \infty$. Hence it is sufficient, if we prove that $(\overline{S_n} - E\overline{S_n})/n \to 0(a.s.)$, if we prove that

$$\sum_k \sigma^2\left(X_k^k\right)/k^2 < \infty.$$

Now

$$\sigma^2\left(X_k^k\right) \le E\left(X_k^k\right)^2 = \int_{-k}^k x^2 dF(x),$$

$$\sum_{j+1}^\infty \frac{1}{k^2} \le \sum_{j+1}^\infty \frac{1}{k(k-1)} = \frac{1}{j} \le \frac{2}{j+1}, \quad \text{for } j \ge 1.$$

Hence

$$\sum_{1}^{\infty}\left(\sigma^2\left(X_k^k\right)/k^2\right) \le \sum_{k=1}^{\infty}\sum_{j=1}^{k}\left[\int_{j-1}^{j} x^2 dF(x) + \int_{-j}^{-j+1} x^2 dF(x)\right]\frac{1}{k^2},$$

$$=\sum_{j=1}^{\infty}\sum_{k=j}^{\infty}\left[\int_{j-1}^{j} x^2 dF(x) + \int_{-j}^{-j+1} x^2 dF(x)\right]\frac{1}{k^2},$$

$$=\sum_{j=1}^{\infty}\left[\int_{j-1}^{j} x^2 dF(x) + \int_{-j}^{-j+1} x^2 dF(x)\right]\sum_{k=j}^{\infty}\frac{1}{k^2},$$

$$\le 2\int_{-\infty}^{\infty} |x|\, dF(x) < \infty.$$

This completes the proof of the theorem.

8.7.1 Kolmogrov's Theorem for SLLN:

Statement: Let $\{X_n\}$ be a sequence of independent and identically distributed random variables. Then a necessary and sufficient condition that $\overline{X}n \xrightarrow{a.s} \mu,$ finite, is that $E\left(X_n\right)$ exists and is equal to μ.

Proof:

Necessity: Let A_n be the event $\left[\left|X_n\right| \ge n\right], n = 0,1,2...,$ so that $A_0 = \Omega$. Hence X_n are i.i.d and have the same distribution, say, as that of X for all n. Thus,

$P(A_n) = P\left[\left|X\right| \ge n\right], A_n \subset A_{n-1}$ and $A_{n-1} - A_n = \left[\left(n-1\right) \le \left|X\right| < n\right].$

So $(n-1)\left[P(A_{n-1}) - P(A_n)\right] \le E\left[\left|X\right| I_{A_{n-1}-A_n}\right] < n\left[P(A_{n-1}) - P(A_n)\right],$ for all n=1,2,... .

Now summing over n, from 1 to ∞, we get $\sum\limits_{n=1}^{\infty}(n-1)P(A_{n-1}) - \sum\limits_{n=1}^{\infty}(n-1)P(A_n) \le E\left|X\right| < \sum\limits_{n=1}^{\infty} nP(A_{n-1}) - \sum\limits_{n=1}^{\infty} nP(A_n)$ (or)

$\sum\limits_{n=1}^{\infty} P(A_n) \le E\left|X\right| < 1 + \sum\limits_{n=1}^{\infty} P(A_n).$ $\rightarrow$**eq(1).** This is valid for all the random variables of the sequence $\{X_n\}$ since they are i.i.d . *As* $\overline{X}n \xrightarrow{a.s} \mu,$ finite, so $\frac{X_n}{n} = \overline{X}n - \frac{n-1}{n}\overline{X}n{-}1 \xrightarrow{a.s} 0.$ The independence of the random variables X_n implies the independence of the events A_n.

Then, by the Borel-Cantelli Lemma, $\sum\limits_{n=1}^{\infty} P\left[\left|\frac{X_n}{n}\right| \ge 1\right] < \infty,$ i.e, $\sum\limits_{n=1}^{\infty} P\left[\left|X_n\right| \ge n\right] < \infty,$ or $\sum\limits_{n=1}^{\infty} P(A_n) < \infty.$ Therefore from eq$\rightarrow$(1) we have $E\left(\left|X\right|\right) < \infty,$ and from the sufficiency condition it follows the $E(X) = \mu.$

Sufficiency: Define , for n=1,2,… , the variables $X_n{}^*$, which are X_n truncated at n:

$X_n^* = X_n$ for $|X| < n$

$= 0$ for $|X| \geq n,$

and $B_m = [(m-1) \leq |X| < m], A_n = [|X| \geq n]$

so $B_m \cap A_n^c = \phi$ if m>n

$= B_m$ if $\text{m} \leq n.$

Consider $I_{B_m}\left[\sum_{n=1}^{\infty} \frac{X^2}{n^2} I_{A_n}{}^c\right] = \sum_{n=1}^{\infty} \frac{X^2}{n^2} I_{B_m \cap A_n{}^c}$

$$= \sum_{n=m}^{\infty} \frac{X^2}{n^2} I_{B_m}$$

$$< \sum_{n=m}^{\infty} \frac{m^2}{n^2} I_{B_m} = m^2\left[\frac{1}{m^2} + \frac{1}{(m+1)^2} + \ldots\right] I_{B_m}$$

$$< \left[1 + m^2 \int_{m+1}^{\infty} \frac{dy}{(y-1)^2}\right] I_{B_m}$$

$$= \left[1 + m^2 \int_{m}^{\infty} \frac{dy}{y^2}\right] I_{B_m} = (1+m) I_{B_m}$$

$$= [2 + (m-1)] I_{B_m} \leq [2 + |X|] I_{B_m},$$

Since $|X| \geq (m-1)$ inside B_m. Summing over m, we get $\sum_{n=1}^{\infty} \frac{X^2}{n^2} I_{A_n}{}^c \leq 2 + |X|,$

So $\sum_{n=1}^{\infty} \frac{Var(X_n^*)}{n^2} \leq \sum_{n=1}^{\infty} \frac{E(X_n{}^x)^2}{n^2} = E\left[\sum_{n=1}^{\infty} \frac{X^2}{n^2} I_{A_n{}^c}\right] \leq 2 + E(|X|) < \infty$

Since $E(|X|) < \infty$. Hence, X_n^* obeys the SLLN, i.e., $\overline{X}n^* - \sum_{k=1}^{n} E(X_k^*) \Big/ n \xrightarrow{a.s} 0 \qquad \rightarrow (2)$

Since $E\left[X_n^*\right] = E(XI_{A_n{}^c}) \rightarrow E(X) = \mu$, by the dominated convergence theorem,

$\frac{1}{n} \sum_{k=1}^{n} E(X_k^*) \rightarrow \mu, as \ \text{n} \rightarrow \infty. \qquad \rightarrow (3)$

By (2) and (3), we have $\overline{X}_n^{*} \xrightarrow{a.s} \mu$

Next we show that $\{X_n\}$ and $\{X_n^{*}\}$ are equivalent sequences in the sense that $\lim_{N} P\left[X_n \neq X_n^{*} \text{ for some } n \geq N\right] = 0$, and this implies that $\to \{X_n\}$ obeys the SLLN if $\{X_n^{*}\}$ obeys if and the limits are the same. We have $P\left[X_n \neq X_n^{*} \text{ for some } n \geq N\right] \leq \sum_{n=N}^{\infty} P\left[X_n \neq X_n^{*}\right] \to 0$ as $N \to \infty$ because by (1),

$$E(|X|) < \infty \Rightarrow \sum_{n=1}^{\infty} P\left[|X_n| \geq n\right] < \infty .$$

$$\text{i.e.,} \quad \sum_{n=1}^{\infty} P\left[X_n \neq X_n^{*}\right] < \infty .$$

8.8 CONCLUSION:

The **Law of Large Numbers (LLN)** is a fundamental theorem in probability theory that describes how the average of a sequence of independent random variables behaves as the sample size increases. It is crucial in statistical estimation, data analysis, and real-world applications such as finance and quality control.

a) **Weak Law of Large Numbers (WLLN)**: This states that the sample mean converges in probability to the expected value as the sample size increases. It ensures that for large samples, observed averages approximate the true mean with high probability.
b) **Necessary and Sufficient Conditions for WLLN**: These conditions determine when the weak law holds, often involving finite expectations and variance constraints.
c) **Chebyshev's Weak Law of Large Numbers**: Derived using **Chebyshev's inequality**, it shows that a sequence of independent and identically distributed (i.i.d.) random variables with finite variance satisfies WLLN.
d) **Khinchin's Weak Law of Large Numbers**: A more general form of WLLN that only requires the existence of the **expected value** (not necessarily finite variance), making it applicable to a broader class of distributions.
e) **Strong Law of Large Numbers (SLLN)**: This states that the sample mean converges **almost surely** to the expected value, meaning that for almost every outcome, the sample mean equals the true mean in the long run.
f) **Kolmogorov's SLLN for the i.i.d. Case**: This is a special case of the strong law that applies when random variables are independent, identically distributed, and have a finite expected value. It guarantees that the sample mean converges with probability1.

8.9 SELF ASSESSMENT QUESTIONS:

1) Derive Chebyshev's Weak Law of Large Numbers using Chebyshev's inequality.
2) Prove Khinchin's Weak Law of Large Numbers for i.i.d. random variables.
3) Compare and contrast Chebyshev's and Khinchin's WLLN in terms of assumptions and applicability.
4) Show how the necessary and sufficient conditions for WLLN guarantee convergence in probability.

5) Discuss an example where the WLLN holds but the SLLN does not.
6) Prove that Kolmogorov's SLLN holds for i.i.d. random variables with finite expectation.
7) Explain how the Law of Large Numbers is used in real-world applications, such as polling or financial modeling.

8.10 SUGGESTED READING BOOKS:

1. Modern probability theory by B. R. Bhat, Wiley EasternLimited.
2. An introduction to probability theory and mathematical statistics by V. K. Rohatgi, John Wiley.
3. AnOutlineofstatisticstheory-1, by A.M.GOON, M.K. Gupta and B. Das gupta, the World Press Private Limited, Calcutta.
4. The Theory of Probability by B.V. Gnedenko, MIR Publishers, Moscow.
5. Discrete distributions -N.L. Johnson and S. Kotz, John wiley & Sons.
6. ContinuousUnivariatedistributions,vol.1&2N.L.JohnsonandS.Kotz,JohnWiley&Sons.
7. Mathematical Statistics-Parimal Mukopadhyay, New Central Book Agency (P) Ltd., Calcutta.

Dr. Syed Jilani

LESSON- 9
DISCRETE DISTRIBUTIONS-I

OBJECTIVES:

After studying this lesson, students will be able to:

- Understand the Concept of Compound Distributions
- Define and Derive the Compound Binomial Distribution
- Understand the concept of truncation
- Differentiate between left-truncated, right-truncated, and doubly-truncated distributions
- Know in detail about truncated binomial distribution

STRUCTURE

9.1 INTRODUCTION:

The **compound binomial distribution** is an extension of the standard binomial distribution, where the number of trials follows a random distribution instead of being fixed.. By introducing randomness in the number of trials, the compound binomial distribution provides greater flexibility in real-world applications. The **truncated binomial distribution** arises when certain values of a standard binomial distribution are excluded, either from the lower or upper end, or both. This modification is necessary when specific outcomes are impossible or unobservable, such as in quality control, genetics, and survey sampling. The probability mass is redistributed among the remaining values to ensure a valid probability model. Both distributions enhance statistical modeling by accounting for uncertainty and constraints in real-world data.

Consider a random variable X whose distribution depends upon the single parameter θ which instead of being regarded as fixed constant is also a random variable following particular distribution. In this case we say that x has a Compound distribution (or) Compose Distribution.

9.2 COMPOUND BINOMIAL DISTRIBUTION:

Let us suppose that $X_1, X_2, \ldots$ are identically independently distributed Bernoulli variates with

$P(X_i = 1) = p$, then $P(X_i = 0) = 1 - p = q$.

For a fixed n, the random variable X= $X_1 + X_2 + \cdots + X_n$ is a binomial variate with parameters n and p and its probability mass function is given by

$$P(X = r) = \binom{n}{r} p^r q^{n-r} \; ; r = 0,1,2, \ldots n$$

which gives the Probability of " r " successes in ' n ' independent trails with constant probability p of success for each trail. Now, suppose that n instead of being regarded as a fixed constant, is also a random variable following Poisson law with Parameter λ. Then

$$P(n = k) = \frac{e^{-\lambda}, \lambda^k}{k!} \; ; k = 0,1,2, \ldots \ldots \ldots \ldots .$$

In such a case X is said to have compound binomial distribution. Then the joint probability function of X and n is given by

$$p(X = r \cap n = k) = p(n = k) \cdot p(x = r/n = k)$$
$$= \frac{e^{\;\lambda} \cdot \lambda^k}{k!} \binom{k}{r} p^r \cdot q^{k-r}$$

$$[\because P(A \cap B) = P(A)P(B)]$$

$\because$ Probability that $P(x - r \mid n = k)$ is the Probability of r successes in k trails.

$\therefore$ Obviously $r \leq k$

$\Rightarrow k \geqslant r$.

Then marginal Distribution of X is given by

$$P(X - r) - \sum_{k=r}^{\infty} P(X - r/n - k)$$
$$= \frac{e^{-\lambda}\lambda^k}{k!} \sum_{k=r}^{\infty} \binom{k}{x} p^r q^{k-r}$$
$$= \frac{e^{-\lambda p} (\lambda p)^r}{k!}$$

9.3 APPLICATIONS OF COMPOUND BINOMIAL DISTRIBUTION:

Some of the practical situations where we could come across the Compound Binomial Distribution are: Suppose that the probability of insect lying n eggs is given by the Poisson distribution $\frac{e^{-\lambda} \cdot \lambda^n}{n!}$ and the probability of an egg developing is p. Assuming natural independence of x, the probability of a total of survivals is given by the poison distribution with parameter λp. The probability that a Radioactive substance gives off $n\beta$ particles in a unit of time is $P(\lambda), (n = 0,1,2, \ldots.)$.

The probability that a given particle will strike a counter and be registered is p. Then the probability of registering $n\beta$ particles in a unit of time is also $p(\lambda p)$. If the probability of number of hits by lightening during any time interval t is $p(\lambda t)$ and it the probability of its hilting and damaging an individual is P then assuming stochastic independent the total damage dousing time t is p$(\lambda + p)$

$$P(x = r) = \sum_{k=r}^{\infty} P(x = r \mid n = k)$$

$$= \frac{e^{-\lambda} \cdot \lambda^k}{k!} \sum_{k=s}^{\infty} \binom{k}{s} p^s q^{k-s}$$

$$= e^{-\lambda} \cdot p^s \lambda^s (\lambda)^{k-s} q^{k-s} \sum_{k=s}^{\infty} \frac{k!}{s!\,(k-s)!\,k!}$$

$$= e^{-\lambda} p^s \lambda^s (\lambda q)^{k-s} \frac{1}{s!} \sum_{k=s}^{\infty} \frac{1}{(k-s)!}$$

$$= \frac{e^{-\lambda} (\lambda p)^s}{s!} \sum_{k=t}^{\infty} \frac{(\lambda q)^{k-s}}{(k-s)!}$$

$$= \frac{e^{-\lambda} (\lambda p)^s}{s!} e^{\lambda q}.$$

$$= \frac{e^{-\lambda(1-q)} (\lambda P)^S}{S!} = \frac{e^{-\lambda P} (\lambda P)^S}{S!}$$

Mean:

$$E(r) = \sum r \cdot P(r)$$

$$= \sum r \cdot \frac{e^{-\lambda p} \cdot (\lambda p)^r}{r!}$$

$$= \sum \frac{r \cdot e^{-\lambda p} \cdot \lambda^r \cdot p^r}{r(r-1)!} = \sum \frac{e^{-\lambda p} \cdot (\lambda p)^{r-1}}{(r-1)!} (\lambda p)$$

$$= e^{-\lambda p} (\lambda P) \sum \frac{(\lambda P)^{r-1}}{(r-1)!} \Rightarrow e^{-\lambda P} (\lambda P) e^{\lambda p} = \lambda P$$

$$E(r^2) = \sum r^2 \cdot P(x)$$
$$= \sum r(r-1+1)p(r)$$
$$= \sum r(r-1)p(r) + \sum rp(r)$$
$$- \sum r(r-1)\frac{e^{-\lambda p} \cdot (\lambda p)^r}{r!} + \sum r \cdot p(r)$$
$$= \sum \frac{e^{-\lambda p} (\lambda p)^{r-2} (\lambda p)^2}{(r-2)!} + \lambda p$$

$$= (\lambda p)^2 + \lambda p$$
$$V(r) = \lambda P[\lambda P + 1] - (\lambda P)^2$$
$$= \lambda^2 P^2 + \lambda P - \lambda^2 P^2 = \lambda P$$

9.4 TRUNCATION:

In statistics, truncation refers to restricting a probability distribution by cutting off values below or above certain thresholds. The resulting distribution is called a **truncated distribution**.

Let $f(x)$ be any function $x(-\infty,\infty)$ Take $a < b$ $belongs\ to$ $(-\infty,\infty)$ then this $f(x)$ will have a another functional form for x belongs to(a,b). This is called Truncated Function. Suppose $f(x)$ is a probability density function or probability mass function of a random variable then $f(x)$ must satisfy the properties of probability density function namely

(1) $f(x) \geqslant 0$
(ii) $\int_{-\infty}^{\infty} f(x)\,dx = 1$.

Take another function $g(x)$ which lies between $a < x < b$

g(x) =0 if $x < a$ and also it is if x>b
If $f(x)$ is a probability density function then $g(x)$ is also a pdf if $g(x)$ satisfies the properties of p.d.f namely
(i) $g(x) \geqslant 0$
(ii) $\int_a^b g(x)dx = 1$.

Define$\int_a^b f(x)dx = c$ (which does not contain) and we know that $f(x)$ is the pdf

$$\therefore f(x) \geqslant 0 \Rightarrow \int_a^b f(x)dx \geqslant 0$$

Let us define $g(x) = \frac{f(x)}{c}$ it $a \leq x \leq b$

$= 0$ Otherwise

Let us take $\int_{-\alpha}^{\alpha} g(x)dx = \int_{-\infty}^{a} g(x)dx + \int_{a}^{b} g(x)dx + \int_{b}^{\infty} g(x)dx$

$$= \int_{a}^{b} g(x)dx = \int_{a}^{b} \frac{f(x)}{c} dx.$$

$$= \frac{1}{c}\int_{a}^{b} f(x)dx = \frac{1}{c} \cdot c = 1 \left(\because \int_{a}^{b} f(x)dx = c\right)$$

c is called as Normalization Constant for truncation $g(x)$ is satisfying the properties of $\mathrm{Pd}f$and $g(x)$ is also a pdf. Here c is called as Normalizing Constant for Truncation.

i) **Left Truncation:** Some points are deleted on Lett ie. $c = \int_{a}^{\infty} f(x)dx$.
$g(x) = \frac{f(x)}{c}$ is called lett Truncation.

ii) **Right Truncation:** Some points are deleted on right. $c = \int_{a}^{b} f(x)dx; g(x) = \frac{f(x)}{c}$ is called Right Truncation.

iii) **Double Truncation:** Truncated left at a and right at b. $c = \int_{a}^{b} f(x)dx$.
$\therefore g(x) = \frac{f(x)}{c}$ is called Double Truncation.

iv) **Truncation:** let $f(x)$ be the pdf or p m f and $F(x)$ be the distribution function of the random variable. Let $[a, b]$ be a subset of the range of the random variable. It p m f is redefined such that the total probability below a and above b is 0 and the entire probability is distributed$[a, b]$. Then we say that distribution of random variable is truncated left at a and right at b.

In the case of discrete random variable $[a, b]$ could be a set of individual elements and in the case of continuous random variable the set is an interval. The p.d.f of the truncated random variable is given by

$$g(x) = \frac{f(x)}{c}$$

$$g(x) = \frac{f(x)}{f(b) - f(a)}$$

9.5 TRUNCATED BINOMIAL DISTRIBUTION:

The p m f of ordinary binomial distribution

is $P(x) = .^{n}c_{x} \cdot p^{x} \cdot q^{n-x}; x = 0,1, \dots n$
$= 0$ otherwise.

Suppose this distribution is truncated left at r_1 and right at r_2 i.e., $(0,1, \dots r_1)$, points are deleated on left and $r_2 + 1, \dots n$ and points are deleated on right then the p m f ofthe truncated binomial distribution will

$$g(x)=\frac{f(x)}{c};\ c=\sum_{r_1+1}^{n-r_2}\binom{n}{x}p^x q^{n-x}$$

$$=\frac{\binom{n}{x}p^x q^{n-x}}{\sum_{r_1+1}^{n-r_2}\binom{n}{x}p^x q^{n-x}}$$

= 0 otherwise

if $r_1 = r_2 = r\ then$

$$g(x)-\frac{\binom{n}{x}p^x q^{n-x}}{\sum_{r_1+1}^{n-r_2}\binom{n}{x}p^x q^{n-x}}$$

is called Symmetrically truncated Binomial Distribution.
Only one point on left is deleted and no points on the right are deleated. It is called zero truncated Binomial Distribution.

The pmf of truncated Binomial Distribution is

$$g(x)=\frac{\binom{n}{x}p^x q^{n-x}}{\sum_{x=1}^{N}\binom{n}{x}p^x q^{n-x}}=\frac{\binom{n}{x}p^x q^{n-x}}{\left[\sum_{x=0}^{n}\binom{n}{x}p^x q^{n-x}\right]^1-q^n}$$

$$=\frac{\binom{n}{x}p^x q^{n-x}}{1-q^n};\ x=1,2,\dots n$$

$$\mathrm{E}(x)=\sum_{x=1}^{n} xg(x)$$

$$=\sum_{x=1}^{n}\frac{x\cdot\binom{n}{x}p^x\cdot q^{n-x}}{1-q^n}$$

$$=\frac{1}{1-q^n}\sum_{x=0}^{n}x\binom{n}{x}p^x q^{n-x}$$

$$=\frac{1}{1-q^n}\cdot np=\frac{np}{1-q^n}$$

is the mean of zero truncated Binomial Distribution.

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

$$E(x^2) = \sum_{x=1}^{n} x^2 g(x) = \sum_{x=1} \frac{x^2\binom{n}{x}p^x q^{n-x}}{1-q^n}$$

$$= \frac{1}{1-q^n} * \sum_{x=0}^{n} x^2 \cdot \binom{n}{x} p^x \cdot q^{n-x}$$

$$= \frac{1}{1-q^n}\{npq + n^2p^2\}$$

$$\therefore V(x) = \frac{1}{1-q^n}\{npq + n^2p^2\} - \frac{n^2p^2}{(1-q^n)^2}$$

$$= \frac{npq}{1-q^n} + \frac{n^2p^2}{1-q^n}\left\{1 - \frac{1}{1-q^n}\right\}$$

$$= \frac{npq}{1-q^n} + \frac{n^2p^2}{1-q^n}\left\{\frac{1-q^n-1}{1-q^n}\right\}$$

$$= \frac{npq}{1-q^n} - \left.\frac{n^2p^2q^n}{(1-q^n)^2}\right\}$$

M.G.F : $M_x(t) = E(e^{tx}) = \sum_{x=1}^{n} e^{tx} \cdot g(x)$

$$= \sum_{x=1}^{n} \frac{e^{tx} \cdot \binom{n}{x}p^x \cdot q^{n-x}}{1-q^n}$$

$$= \frac{1}{1-q^n}\sum_{n=1}^{n} \binom{n}{x} (pe^t)^x \cdot q^{n-x}$$

$$= \frac{1}{1-q^n}\left[\sum_{x=0}^{n} \binom{n}{x} (pe^t)^x q^{n-x} - q^n\right]$$

$$- \frac{1}{1-q^n}[(q+pe^t)^n - q^n]$$

$= \frac{(q+pe^t)^n}{1-q^n} - \frac{q^n}{1-q^n}$ is

the m g f of zero truncated Binomial Distribution-Similarly, the characteristic function of zero truncated Binomial Distribution is

$$\frac{(q+pe^{it})^n}{1-q^n} - \frac{q^n}{1-q^n}$$

Additive property does not hold because zero is deleted and the range contains only positive integers. In that sense zero truncated Binomial Distribution is sometimes called a Positive Binomial Distribution.

9.6 CONCLUSION:

The **compound binomial distribution** extends the binomial model by allowing the number of trials to be random, making it useful in insurance, biology, and network analysis. The **truncated binomial distribution** excludes certain outcomes, adjusting probabilities accordingly. It is applied in quality control, genetics, and sampling where specific values are unobservable.

9.7 SELF ASSESSMENT QUESTIONS:

1. How does a compound binomial distribution differ from a standard binomial distribution, and in what scenarios is it useful?
2. What is the impact of introducing randomness in the number of trials on the properties of a compound binomial distribution?
3. What are the different types of truncation in a binomial distribution, and how do they affect probability calculations?
4. Why is a truncated binomial distribution necessary in real-world applications, and how is it used in quality control and genetics?
5. How do compound binomial and truncated binomial distributions help in handling over dispersion and constrained data in statistical modeling?

9.8 SUGGESTED BOOKS FOR FURTHER READING:

1. "Discrete Distributions" – Norman L. Johnson, Adrienne W. Kemp, Samuel Kotz (Wiley)
2. "Univariate Discrete Distributions" – Norman L. Johnson, Adrienne W. Kemp, Samuel Kotz (Wiley)
3. "Statistical Distributions" – Merran Evans, Nicholas Hastings, Brian Peacock (Wiley)
4. "Statistical Size Distributions in Economics and Actuarial Sciences" – Christian Kleiber, Samuel Kotz (Wiley)
5. "Discrete Statistical Distributions" – T.P. Hutchinson, C.D. Lai (Routledge)

Dr. I.B.N. Hima Bindu

LESSON 10
DISCRETE DISTRIBUTIONS -II

OBJECTIVES:

After studying this lesson, students will be able to:

- Explain the concepts of compound Poisson and truncated Poisson distributions and how they extend the standard Poisson model.
- Derive and interpret key properties such as the mean, variance, and probability mass function for both distributions.
- Understand the impact of truncation and compounding on probability calculations and distributional behavior.
- Compare these distributions with standard Poisson and other related models, highlighting their advantages in real-world applications.
- Apply compound Poisson models and truncated Poisson models to real time problems.

STRUCTURE:

10.1 INTRODUCTION:

A **compound Poisson distribution** is an extension of the standard Poisson distribution, where each event contributes a random value instead of a fixed unit. It is commonly used in risk modeling, queuing theory, and insurance, where the number of occurrences follows a Poisson process, but the impact of each event varies. This distribution helps model situations where both the frequency and severity of events are uncertain, such as the total claim amounts in insurance or packet arrivals in network traffic.

A **truncated Poisson distribution** is a modified version of the Poisson distribution in which certain values are excluded. Truncation can be at the lower end, upper end, or both, depending on the problem context. For example, in biological studies, data may be truncated because very small counts are not observed, or in accident data analysis, certain extreme values may be removed. The distribution is adjusted to ensure valid probability calculations for the remaining outcomes.

10.2 COMPOUND POISSON DISTRIBUTION:

Let x be a $P(\lambda)$ so that

$$p(x=r)=\frac{e^{-\lambda}\lambda^{\gamma}}{r!}\ ;\ r=0,1,2,\ldots$$

Where λ itself is a continuous random variable with generalized gamma density

$$g(\lambda)=\begin{cases}\dfrac{a^{\gamma}}{\Gamma(v)}e^{-a\lambda}\lambda^{\gamma-1}; \lambda>0, a>0, \gamma>0\\ 0, \lambda\leq 0\end{cases}$$

Let us consider the two dimensional random vector (x,λ) in which one variable is discreate and the other is continuous. For a constant h>0 and $\lambda_1>0$, the joint density of x and λ is given by

$$P(\mathrm{X}=r\cap\lambda_1\leqslant\lambda\leqslant\lambda_1+h)=P(\lambda_1\leqslant\lambda\leqslant\lambda_1+h)P(x=r\mid\lambda_1\leqslant\lambda\leqslant\lambda_1+h)$$

Dividing both sides by h and proceeding to the limits as $h\to 0$ we get

$$\lim_{h\to 0}\frac{p(x=\gamma\cap\lambda_1\leq\lambda\leq\lambda_1+h)}{h}=\lim_{h\to 0}p(x=r/\lambda_1\leqslant\lambda<\lambda_1+h)$$
$$\times\lim_{h\to 0}\frac{p(\lambda_1\leq\lambda\leq\lambda_1+h)}{h}$$

But

$$\lim_{h\to 0}\frac{p(\lambda_1\leq\lambda\leq\lambda_1+h)}{h}=\lim_{h\to 0}\frac{G(\lambda_1+h)-G(\lambda_1)}{h}$$
$$=G^1(\lambda_1)=g(\lambda_1)$$

Where $G(.)$ *is the distribution function and* $g(.)$ *is the* Pdf.

$$\therefore\lim_{h\to 0}\frac{p(x=\gamma\cap\lambda_1\leqslant\lambda\leqslant\lambda_1+h)}{h}=\frac{e^{-\lambda_1}\lambda_1^{\gamma}}{r!}.$$
$$\frac{a^{\gamma}}{r(\lambda)}\lambda_1^{\gamma-1}\cdot e^{-a\lambda_1}$$

Integrating w. r. to λ_1 over o to ∞ and using gamma integral, the marginal probability function of x is given by

$$P(x=r)=\frac{a^{\gamma}}{\Gamma(v)r!}\int_0^{\alpha} e^{-(1+a)\lambda}\lambda^{\gamma+v-1}d\lambda$$
$$=\frac{a^{\gamma}}{\Gamma(v)\gamma!}\cdot\frac{\Gamma(\gamma+\gamma)}{(1+a)^{\gamma+\gamma}}$$
$$=\left(\frac{a}{1+a}\right)^{\gamma}\frac{\gamma(\gamma+1)(\gamma+2)+\cdots(v+\gamma-1)}{(1+a)^{\gamma}\gamma!}$$
$$=\left(\frac{a}{1+a}\right)^{\gamma}(-1)^{\gamma}\binom{-\gamma}{\gamma}\left(\frac{1}{1+a}\right)^{\gamma}$$
$$=\binom{v}{\gamma}p^{\gamma}(-q)^{\gamma};r=0,1,2\cdots$$

where $p=\frac{a}{1+a}$; $q=1-p=\frac{1}{1+a}$.
Thus the marginal distribution of x is a negative binomial with parameters (γ,p).

10.3 COMMON USE CASES:

The **compound Poisson distribution** is widely used in real-world applications where both the number of events and their individual impacts vary. Some key use cases are:

1. **Insurance and Risk Management**
   - Models total claim amounts, where the number of claims follows a Poisson process, and each claim amount is a random variable.
   - Used in actuarial science to estimate risk and determine insurance premiums.
2. **Finance and Economics**
   - Describes stock market fluctuations where the number of transactions follows a Poisson process, and each transaction has a variable impact.
   - Models aggregate losses in credit risk analysis.
3. **Telecommunications and Network Traffic**
   - Used to model data packet arrivals in communication networks, where packets arrive randomly, and their sizes vary.
   - Helps in network capacity planning and performance optimization.
4. **Reliability and Maintenance**
   - Models equipment failures where the number of failures follows a Poisson process, and the severity of each failure is random.
   - Used in predictive maintenance and reliability engineering.
5. **Biology and Epidemiology**
   - Applied in genetics to model the occurrence and severity of mutations.
   - Used in epidemiology to study disease outbreaks, where the number of cases follows a Poisson process, and the severity varies.

10.4 TRUNCATED POISSON DISTRIBUTION:

The p m f of the ordinary Poisson Distribution is.

$$f(x)=\frac{e^{-\lambda}\lambda^{x}}{x!}\quad ; x=0,1,2,\ldots\ldots\ldots\infty$$

$$=0\ ;\ \text{otherwise}$$

Suppose this Distribution is truncated left at r_1 and right at r_2 i.e., $0,1,2,\ldots r_1$. points are deleted on left and $r_{2+1},\ldots\ldots\infty$ points are deleted on right then the p m f of truncated Poisson Distribution will be

$$g(x)=\frac{f(x)}{c};\ \text{Where } c=\sum_{r_1+1}^{r_2}\frac{e^{-\lambda}\cdot\lambda^x}{x!}$$
$$=g(x)=\frac{e^{-\lambda}\cdot\lambda^x/x!}{\sum_{r_1+1}^{r_2}\frac{e^{-\lambda}\cdot\lambda^x}{x!}}$$
$$=0\text{ ; otherwise.}$$

is Called the truncated Poisson Distribution truncated left at r_1 and right at r_2.

Zero Truncated Poisson Distribution:

If only one point on left is deleted i. e, called zero Truncated Poisson Distribution.

$$g(x)=\frac{f(x)}{c}\ \text{Where } c\cdot\sum_{1}^{\alpha}\frac{e^{-\lambda}\lambda^x}{x!}$$
$$=\frac{\frac{e^{-\lambda}\cdot\lambda^x}{x!}}{\sum_{x=0}^{\infty}\frac{e^{-\lambda}\cdot\lambda^x}{x!}-e^{-\lambda}}=\frac{\frac{e^{-\lambda}\lambda^x}{x!}}{1-e^{-\lambda}}$$

Since $\sum_{x=0}^{\infty}\frac{e^{-\lambda}\lambda^x}{x!}=1$ Distribution function $g(x)=\frac{\frac{e^{-\lambda}\lambda^x}{x!}}{1-e^{-\lambda}}$

$=0$ otherwise.

Moment Generating Function, Mean and Variance

$$\text{M.G.f}: M_x(t)=E(e^{tx})=\sum_{x=1}^{\infty}e^{tx}g(x)$$
$$-\sum_{x=1}^{\infty}e^{tx}\cdot\frac{\frac{e^{-\lambda}\lambda^x}{x!}}{1-e^{-\lambda}}$$
$$=\frac{1}{1-e^{-\lambda}}\sum_{x=1}^{\infty}e^{tx}\cdot\frac{e^{-\lambda}\cdot\lambda^x}{x!}$$
$$=\frac{1}{1-e^{-\lambda}}\sum_{x=1}^{\infty}\frac{(\lambda e^t)^x\cdot e^{-\lambda}}{x!}$$

$$= \frac{1}{1-e^{\lambda}}\left\{\sum_{x=1}^{\infty}\frac{(\lambda e)^{x}e^{\lambda}}{x!} - e^{-x}\right\}$$

$$= \frac{1}{1-e^{\lambda}}\left\{e^{-\lambda}\sum_{x=0}^{\infty}, \frac{(\lambda e^{t})^{x}}{x!} - 1\right)\right\}\left(\because \frac{k^{x}}{x!} - e^{k}\right)$$

$$= \frac{e^{-\lambda}}{1-e^{-\lambda}}\{e^{\lambda e^{t}-1} = 1\}$$

Mean

$$M_x'(t) = \frac{e^{-\lambda}}{1-e^{-\lambda}}\left[e^{\lambda e^{t}}\lambda e^{t}\right]_{t=0}\Big)$$

$$= \frac{e^{-\lambda}}{1-e^{-\lambda}}[e^{\lambda}\lambda]$$

$$= \frac{\lambda}{1-e^{-\lambda}}$$

$$E(x) = \sum_{x=1}^{\infty} xg(x)$$

$$= \sum_{x=1}^{\alpha} x\frac{\frac{e^{-\lambda}\lambda^{x}}{x!}}{1-e^{-\lambda}}$$

$$= \frac{1}{1-e^{-\lambda}}\sum_{x=0}^{\alpha}\frac{x\cdot e^{\lambda}\lambda^{x}}{x!}\left(\sum_{x=0}^{\alpha} x\frac{e^{\lambda}\cdot\lambda^{x}}{x!} = \lambda\right)$$

$$= \frac{\lambda}{1-e^{-\lambda}}$$

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

$$E(x^{2}) = \sum_{=1}^{\infty} x^{2}g(x)$$

$$= \sum_{x=1}^{\infty} x^2 \frac{\frac{e^{\lambda}\lambda^x}{x!}}{1-e^{-\lambda}}$$

$$= \frac{1}{1-e^{-\lambda}} \sum_{x=1}^{x} x^2 \cdot \frac{e^{-\lambda} \cdot x^x}{x!}$$

$$= \frac{1}{1-e^{-\lambda}} \sum_{x=1}^{\infty} (x(x-1)+x) - \frac{e^{-\lambda}\lambda^x}{x!}$$

$$= \frac{1}{1-e^{-\lambda}} \left\{ \sum_{x=1}^{\infty} \left[\frac{x(x-1)\lambda^x e^{-\lambda}}{x!} + \frac{xe^{-\lambda} \cdot \lambda^x}{x!} \right] \right\}$$

$$= \frac{1}{1-e^{-\lambda}} \left[\sum_{x=0}^{\infty} \frac{x(x-1)e^{-\lambda}\lambda^x}{x!} + \sum_{x=0}^{\infty} \frac{x \cdot e^{-\lambda}\lambda^x}{x!} \right]$$

$$= \frac{1}{1-e^{-\lambda}} \left[\sum_{x=\lambda}^{\infty} \frac{e^{\lambda}\lambda^2\lambda^{x-2}}{(x-2)!} \right] + \lambda.$$

$$= \lambda^2 + \lambda/1 - e^{\lambda} = \frac{\lambda(\lambda+1)}{1-e^{-\lambda}}$$

$$E(x^2) = \frac{\lambda e^{-\lambda}}{1-e^{\lambda}} \left[e^t e^{\lambda e^t} \lambda e^t + e^{\lambda e^t} e^t \right] \text{ (OR)}$$

$$= \frac{\lambda e^{-\lambda}}{1-e^{-\lambda}} \left[e^{\lambda}\lambda + e^{\lambda} \right]$$

$$= \frac{\lambda(\lambda+1)}{1-e^{-\lambda}}.$$

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

$$\frac{\lambda(\lambda+1)}{1-e^{-\lambda}} - \frac{\lambda^2}{(1-e^{-\lambda})^2}$$

$$= \frac{\lambda}{1-e^{-\lambda}} \left[(\lambda+1) - \frac{\lambda}{1-e^{-\lambda}} \right]$$

$$= \frac{\lambda}{1-e^{-\lambda}} \left| \frac{\lambda + 1 - \lambda e^{-\lambda} - e^{-\lambda} - \lambda}{1-e^{-\lambda}} \right|$$

$$= \frac{\lambda}{(1-e^{-\lambda})^2} \left[1 - \lambda e^{-\lambda} - e^{-\lambda} \right]$$

Mean and variance are not equal.

$$E(x^3) = \frac{\lambda(\lambda+1)(\lambda+2)}{1-e^{-\lambda}}$$

$$(E(X^4) = \frac{\lambda(\lambda+1)(\lambda+2)(\lambda+3)}{1-e^{-\lambda}}$$

10.5 COMMON USE CASES:

The truncated Poisson distribution is a modified version of the standard Poisson distribution where certain values (typically low or high) are excluded. It is useful in scenarios where observations are censored or restricted. Below are some common applications:

1. Accident and Insurance Data Analysis
   - Used when minor accidents or claims below a certain threshold are not reported.
   - Helps in modeling only significant claims while ignoring small-value claims.
2. Biological and Ecological Studies
   - Applied in species abundance models where very rare species may not be recorded.
   - Used in population studies where data collection is limited to a specific range.
3. Quality Control and Defect Analysis
   - Models the number of defective items in a batch when no-defect items are not recorded.
   - Used in industrial inspection processes where only high-defect products are considered.
4. Queueing Systems and Service Industry
   - Helps in analyzing waiting times where very short or very long queues are not recorded.
   - Used in modeling call center operations where extremely short calls are ignored.
5. Medical and Epidemiological Studies
   - Used in disease incidence models where mild or undiagnosed cases are not included.
   - Applied in survival analysis where only patients with a minimum number of hospital visits are considered.

10.6 CONCLUSION:

A compound Poisson distribution generalizes the Poisson model by allowing each event to contribute a random value instead of a fixed unit. It is useful in risk modeling, insurance claims, and network traffic analysis, where both event counts and magnitudes vary. A truncated Poisson distribution modifies the standard Poisson model by excluding certain values, either from the left, right, or both. It is applied in accident analysis, quality control, and ecological studies, where small or extreme counts are unobservable or irrelevant. Both distributions help create more realistic statistical models for practical applications in diverse fields.

10.7 SELF ASSESSMENT QUESTIONS:

1. How does a compound Poisson distribution differ from a standard Poisson distribution, and in what scenarios is it useful?
2. What are the key properties of a compound Poisson distribution, and how does the introduction of random event sizes affect its mean and variance?
3. What is a truncated Poisson distribution, and how is the probability mass adjusted when certain values are excluded?

4. In which real-world applications would a truncated Poisson distribution be more appropriate than a standard Poisson model?
5. How do compound Poisson and truncated Poisson distributions refine statistical modeling in areas like insurance, quality control, and risk assessment?

10.8 SUGGESTED BOOKS FOR FURTHER READING:

1. "Univariate Discrete Distributions" – Norman L. Johnson, Adrienne W. Kemp, Samuel Kotz (Wiley)
2. "Discrete Distributions" – Norman L. Johnson, Adrienne W. Kemp, Samuel Kotz (Wiley)
3. "Statistical Size Distributions in Economics and Actuarial Sciences" – Christian Kleiber, Samuel Kotz (Wiley)
4. "Stochastic Processes" – Sheldon M. Ross (Wiley)
5. "Probability and Statistical Inference" – Robert V. Hogg, Elliot A. Tanis, Dale L. Zimmerman (Pearson)

Dr. I.B.N. Hima Bindu

LESSON -11

CONTINUOUS DISTRIBUTIONS (LAPLACE, WEIBULL, LOGISTIC AND PARETO DISTRIBUTIONS)

OBJECTIVES:

After studying this topic, students will be able to:

- Understand the Characteristics of Continuous Distributions
- Analyze Probability Density and Cumulative Distribution Functions
- Compute Moments and Statistical Properties
- Compare and Contrast Different Continuous Distributions
- Apply These Distributions in Practical Scenarios

STRUCTURE:

11.1 INTRODUCTION:

Continuous probability distributions play a crucial role in statistical modeling and data analysis. Among them, the **Laplace, Weibull, Logistic, and Pareto distributions** are widely used in various fields due to their unique properties and applications. The **Laplace distribution** is symmetric and exhibits heavier tails than the normal distribution, making it suitable for modeling financial returns, signal processing, and Bayesian statistics. The **Weibull distribution** is a versatile model frequently used in reliability analysis, survival studies, and failure time predictions. The **Logistic distribution** resembles the normal distribution but has heavier tails and a closed-form cumulative distribution function. The **Pareto distribution** is a heavy-tailed distribution commonly applied in economics, finance, and risk analysis. It effectively models wealth distribution, insurance claims, and internet traffic data, where a small percentage of values contribute to most of the total impact. Understanding these distributions and their properties allows for better data interpretation and decision-making across various domains.

11.2 LAPLACE DISTRTBUTION (or Double Exponential Distribution):

Consider the standard Exponential namely e^{-x} where range is $(0\ \infty)$ then the range can be extended to negative part namely $(-\infty, 0)$ also by giving a new density as follows.

Let $f(x) = \frac{1}{2}e^{-|x|}\ \ -\infty < x < \infty$
Then $f(x)$ satisfies the two conditions

(i) $f(x) \geqslant 0$
(ii) $\int_{-\infty}^{\infty} f(x)dx = 1.$

This Density function is called standard Double Exponential Density (or) Laplace Density or it also known as Laplace second law of errors invented by Laplace in 1774

To Prove that $\int_{-\infty}^{\infty} f(x)dx = 1.$

Proof:

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{\infty} \frac{1}{2}e^{-|x|}dx$$

$$= \frac{1}{2}\int_{-\infty}^{\infty} e^{-|x|}dx$$

Consider

$$= \frac{1}{2}\left[\int_{-\infty}^{0} e^{-(x)}dx + \int_{0}^{\infty} e^{-(x)}\,dx\right]$$

$$= \frac{1}{2}\int_{-\infty}^{0} e^{-(-x)}dx + \frac{1}{2}\int_{0}^{\infty} e^{-x}dx.$$

$$= \frac{1}{2}\int_{-\infty}^{0} e^{x}\,dx + \frac{1}{2}\int_{0}^{\infty} e^{-x}dx.$$

since $|x| = x\ \ if x \geqslant 0$

$|x| = -x$ if $x < 0.$

put $x = \quad \mu$ in the first integral and $dx = \quad du.$
$x = v$ in the second integral.

$$= \frac{1}{2}\int_{-\infty}^{0} e^{-u}(du) + \frac{1}{2}\int_{0}^{\infty} e^{-v}dv$$

$$= \frac{1}{2}\int_{0}^{\infty} e^{u}du + \frac{1}{2}\int_{0}^{\infty} e^{-v}\,dv.$$

$$- \frac{1}{2}\cdot 1 + \frac{1}{2}\cdot 1 - 1.$$

Hence proved.

Further we have $f(-x) = \frac{1}{2}e^{-|-x|} = \frac{-1}{2}e^{-|x|} = f(x)$

$(\because |-x| = |x|)$

$f(-x) = f(x) \Rightarrow f(x)$ is symmetric about zero,

11.2.1 Characteristic Function:

$$E\left(e^{itx}\right)=\int_{-\infty}^{\infty} e^{itx}\cdot f(x)dx$$
$$=\frac{1}{2}\int_{\infty}^{\infty} e^{itx}\, e^{-|x|}dx$$
$$=\frac{1}{2}\int_{-\infty}^{0} e^{itx}e^{x}dx+\frac{1}{2}\int_{0}^{\infty} e^{itx}e^{-x}dx$$

$=\frac{1}{2}\int_0^\infty e^{itx}\cdot e^{-x}dx$ is the characteristic furnation of Exponential Distribution.

Put $x=-\mu$
$\Rightarrow dx=-d\mu$ in first integral
$x=\text{v}$ in the second integral.

$dx=dv$

$$=\frac{1}{2}\int_{-\infty}^{0}-e^{-itu}e^{-u}du+\frac{1}{2}\int_{0}^{\infty} e^{-itv}\cdot e^{-v}dv$$
$$\left(\because \int_{0}^{\infty} e^{itx}e^{-x}dx=(1-it)^{-1}\right)^{1}$$

The second integral is the characteristic function of standard Negative Exponential is given by $(1-it)^{-1}$.

The first integral is almost same as the second integral except that i is replaced by –i which is the complex conjugate of the second integral.

$$\therefore cf=\frac{1}{2}(1+it)^{-1}+\frac{1}{2}(1-it)^{-1}$$
$$=\frac{(1+it)^{-1}+(1-it)^{-1}}{2}$$

we have $(1+t)^{-1}=1-t+t^2-t^3+t^4-$................

$(1-t)^{-1}=1+t+t^2+t^3+t^4+\cdots$

On adding $(1+t)^{-1}+(1-t)^{-1}=2(1+t^2+t^4+\cdots)$
similarly $c\cdot f=\frac{1}{2}(1+it)^{-1}+\frac{1}{2}(1-it)^{-1}$

$$=\frac{(1+it)^{-1}+(1-it)^{-1}}{2}$$

$=1+(it)^2+(it)^4+\cdots$

$=1+i^2t^2+i^4t^4t-$

$=1+i^2t^2+i^4t^4+\cdots$
$=1-(t^2)^1+(t^2)^2-(t^2)^3+\cdots$ Where $t=t^2$.

$=1-T+T^2-T^3+\cdots$

$= (1+T)^{-1}$

$= (1+t^2)^{-1}$

$= \dfrac{1}{1+t^2}$

Hence the characteristic function for of Laplace distribution is $\frac{1}{1+t^2}$

$= 1 - t^2 + t^4 - t^6 + \cdots$

11.2.2 Moments:

coefficient of $\frac{(it)^k}{k!}$ in the characteristic function gives kth moment..

$E(x^k) = 0$ if k is odd

$\quad = k!$ if k is Even.

$F(x^2) = 2!$

Variance $V(x) = E(x^2) - [E(x)]^2$

$= 2! - 0$
$= 2.$

∴ Moments are

$$E(x) - 0$$
$$E(x^2) = 2!$$
$$E(x^3) = 0 = \mu_3$$
$$E(x^4) = 4! = \mu_4$$
$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = 0$$
$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{4!}{2^2!} = \frac{4 \times 3 \times 2 \times 1}{2 \times 2}$$
$$= 6$$

11.2.3 Moment Generating Function $M_x(t)$:

$$M_x(t) = F(e^{tx}) = \int_{-\infty}^{\infty} e^{tx}\frac{1}{2}e^{-|x|}dx$$

$$= \frac{1}{2}\int_{-\infty}^{0} e^{tx}e^{x}dx + \int_{0}^{\infty} e^{tx}e^{-x}dx$$

$$= \frac{1}{2}\Big|_{-\infty}^{0} e^{(t+1)x}dx + \frac{1}{2}\int_{0}^{\infty} e^{-(1-t)x}dx$$

$$= \frac{1}{2}\int_{-\infty}^{0} d\left[\frac{e^{(1+t)x}}{1+t}\right] + \frac{1}{2}\int_{0}^{\infty} d\left[\frac{e^{-(1-t)x}}{-(1-t)}\right]$$

$$= \frac{1}{2}\left[\frac{1}{1+t}\left[e^{(1+t)x}\right]_{-\infty}^{0} - \frac{1}{(1-t)}\left[e^{-(1-t)x}\right]_{0}^{\infty}\right.$$

$$= \frac{1}{2}\left[\frac{1}{1+t}(1-0) - \frac{1}{1-t}(0-1)\right]$$

$$= \frac{1}{2}\left[\frac{1-t+1+t}{(1-t)(1+t)}\right] = \frac{2}{2(1-t^2)} = \frac{1}{(1-t^2)}$$

11.2.4 Cumulant Generating Function (CGF):

$$k_x(t) - \log M_X(t)$$

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

$$= \log 1 - \log(1-t^2)$$

$$= 1 + t^2 + \frac{t^4}{4} + \cdots$$

$k_r = 0$ if x is odd

$k_r = 2\lfloor(r-1)!\rfloor$ if r is even

11.3 TWO PARAMETERS LAPLACE DISTRIBUTION:

If μ is location Parameter, σ is Sale Parameter then pdf is given by

$$f(x) = \frac{1}{2\sigma}e^{-\left|\frac{x-\mu}{\sigma}\right|}; \qquad -\infty < x < \infty, -\infty < \mu < \infty, \sigma > 0$$

=0 otherwise

Properties:

$E(X) = \mu$

$V(X) = 2\sigma^2$

S.D of $X = \sigma\sqrt{2}$

$\beta_1 = 0;\ \beta_2 = 6$

$$\text{Mean Deviation} = \int_{-\infty}^{\infty} |x| \frac{1}{2} e^{-|x|} dx$$

$$= \int_{-\infty}^{0} -x \frac{1}{2} e^{x} dx + \int_{0}^{\infty} x \frac{1}{2} e^{-x} dx$$

Put $x = -\mu$

$$\Rightarrow dx = -d\mu$$

$$= \frac{1}{2} \int_{0}^{\infty} \mu e^{-\mu} d\mu + \frac{1}{2} \int_{0}^{\infty} \mu e^{-\mu} d\mu$$

$$= \frac{1}{2} + \frac{1}{2} = 1 \qquad \therefore \left(\int_{0}^{\infty} \mu e^{-\mu} d\mu = 1 \right)$$

$$\text{Variance} = 2.$$

$$SD = \sqrt{2}.$$

$$\therefore \frac{\text{Mean Deviation}}{\text{Standard Deviation}} = \frac{1}{\sqrt{2}}$$

11.4 WEIBULL DISTRIBUTION:

Take a constant $c > 0$ and a r.v X^c follows Exponential (standard negative exponential) distribution then X follows Weibull Distribution.

$\log x \sim$ Normal Dist

then we say $x \sim$ log. Normal.

$$y = x^c;\ dy = cx^{c-1} dx$$

We know that y follows Exponential

ie. $f(y) = e^{-y} dy;\ y = x^c; c > 0, x > 0$

then $f(x) = c \cdot x^{c-1} e^{-x^c} dx$,which is the pdf of Weibull Distribution.

Definition: A continuous r .v X assuming non-negative values is said to have Weibull Distribution with parameter $c(c > 0)$ and if it's pdf is given by $f(x) = cx^{c-1} e^{-x^c}; x > 0, c > 0$

$= 0$ otherwise

Weibull Distribution was Discovered by Swedish mathematician Weibull in 1939 cumulative. Distribution for $F(x)$ is given by

$$F(x) = \int_{-\infty}^{x} f(x) dx = \int_{-\infty}^{0} f(x) dx + \int_{0}^{\infty} f(x)\, dx$$

$$= \int_{0}^{x} f(x) dx = \int_{0}^{x} e^{-x^c} cx^{c-1} dx$$

Put $x^c = y \Rightarrow dy = cx^{c-1} dx$

$$\int_0^x e^{-x^c} c x^{c-1} dx$$
$$= \int_0^x e^{-y} dy = \left[\frac{e^{-y}}{-1}\right]_0^{x^c}$$
$$= \frac{e^{-x^c}}{-1} + 1 = 1 - e^{-x^c}$$

here C is Called Shape Parameter (or) Power Parameter.

11.5 THREE PARAMETERS WEIBULL DISTRIBUTION:

let $z - \left(\frac{x-\mu}{\sigma}\right)^c$ Where $-\infty < \mu < \infty$

$\sigma > 0, c > 0$

$x \sim$ Standard Exponential

$$\text{i.e, } x \sim e^{-z} dz$$
$$\therefore \text{dz} = c\left(\frac{x-\mu}{\sigma}\right)^{c-1} \frac{1}{\sigma} dx$$

$\therefore$ P.d.f is given by

$$f(x) = e^{-\left(\frac{x-\mu}{\sigma}\right)^c} \left(\frac{x-\mu}{\sigma}\right)^{c-1} c \frac{1}{\sigma} dx$$
$$= \frac{c}{\sigma^c} (x-\mu)^{c-1} e^{-\left(\frac{x-\mu}{\sigma}\right)^c},$$
$$-\infty < x < \infty$$
$$-\infty < \mu < \infty$$
$$c > 0, \sigma > 0$$

This is called Three Parameters μ, σ, c weibull distribution. Here μ is called location parameter, σ is called scale parameter and c is called shape (or) Power parameter

Note: If we take $\mu = 0, \sigma = 1$ we get the pdf to of Standard Weibull Distribution

11.5.1 Moments:

$$\mu_r^1 = E(x^r) = \int_0^\infty x^r \cdot f(x)dx$$
$$= \int_0^\infty x^r e^{-x^c} c x^{c-1} dx$$
$$= \int_0^\infty x^r (c x^{c-1} dx) e^{-x^c} \qquad x^c = z$$
$$= \int_0^\infty z^{r/c} e^{-z} dz \qquad \Rightarrow x = z^{1/c}$$
$$= \int_0^\alpha z^{\left(\frac{x}{c}+1\right)-1} e^{-z} dz \qquad x^x = z^{r/c}$$
$$= \Gamma\left(1 + \frac{r}{c}\right) \qquad c x^{c-1} dx = dz$$
$$E(x^{ri}) = \Gamma\left(1 + \frac{r}{c}\right) \qquad \left(\frac{x}{c} + 1 = n \; say\right)$$
$$\left(\therefore \int_0^\infty e^{-x} e^{n-x} dx = \Gamma \mathrm{n}\right)$$

It can be obtained from tables of Crammer function If $r = 1,2,3,4$ then we get the four raw moments

11.5.2 Properties:

$$\mu_1' = E(x) = \Gamma\left(1 + \frac{1}{c}\right)$$
$$\mu_2' - E(x^2) - \Gamma\left(1 + \frac{2}{c}\right)$$
$$\mu_2 = \mu_2' - (\mu_1')^2$$
$$= \Gamma\left(1 + \frac{2}{c}\right) - \left[\Gamma\left(1 + \frac{1}{c}\right)\right]^2$$

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

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

If $c = 3.6$ β_1 is very small number 0 . (Curve of β_1 is bell shape and tends to normal ie., if $c = 3.6$ Weibull follows Normal)

11.5.3 Moment Generating Function: We know that $\frac{t^r}{r!}$ coefficient is r th moment of x

$$\therefore MGF = \sum_{r=0}^{\infty} \frac{t^r}{r!} E(x^r) \Bigg| E(e^{tx}) = E\left[\sum_{r=0}^{\infty} \frac{(t-x)^r}{r!}\right]$$
$$= \sum_{r=0}^{\infty} \frac{t^r E(x^r)}{r!}$$

$$\sum_{r=0}^{\infty} \frac{t^r}{r!}\Gamma\left(1+\frac{r}{c}\right)$$

11.5.4 Characteristic function: $\emptyset_x(t) = \sum_{x=0}^{\infty} \frac{(it)^r}{r!}\Gamma\left(1+\frac{r}{c}\right)$ Additive property does not hold good.

If c_1, c_2 are two shape parameters of two independent Weibull r .v's say X_1, X_2

$$\text{c.f of } x_1 + x_2 = \sum_{r=0}^{\infty} \frac{(it)^r}{r!}\Gamma\left(1+\frac{x}{c_1}\right)\sum_{r=0}^{\infty} \frac{(it)^r}{r!}\Gamma\left(1+\frac{x}{c_2}\right)$$

$$\neq \sum_{r=0}^{\alpha} \frac{(it)^r}{r!}\Gamma\left(1+\frac{r}{c_1+c_2}\right)$$

11.6 LOGISTIC DISTRIBUTION:

The **logistic distribution** is a continuous probability distribution widely used in statistics, machine learning, and reliability analysis. It resembles the normal distribution but has heavier tails, making it useful in modeling growth processes, logistic regression, and survival analysis.

11.6.1 Probability Density Function (p.d.f.):

The probability density function (p.d.f.) of the logistic distribution is given by:

$$f(x) = \frac{e^{\frac{-(x-\mu)}{\delta}}}{\delta\left(1+e^{\frac{-(x-\mu)}{\delta}}\right)^2} \quad -\infty < x < \infty$$

where:

μ is the location parameter (mean and median).

- δ is the scale parameter, which determines the spread of the distribution.

The shape of the logistic distribution is similar to a normal distribution but with a sharper peak and heavier tails.

11.6.2 Cumulative Distribution Function (c.d.f.):

The cumulative distribution function (c.d.f.) is given by:

$$F(x) = \frac{1}{1+c^{-\frac{x-\mu}{\delta}}} - \infty < x < \infty$$

This function has a **sigmoidal (S-shaped) curve**, making it widely used in logistic regression and neural networks.

Mean, Variance, and Moments:

- Mean: E[X]=μ
- Variance: Var(X)=$\frac{\pi^2 \delta^2}{3}$
- Skewness: Zero (symmetrical distribution)
- Kurtosis**:** Higher than the normal distribution, making it more prone to extreme values.

11.6.3 Properties:

1. Symmetry – The distribution is symmetric about μ\muμ, similar to the normal distribution.
2. Heavy Tails – The probability of extreme values is higher compared to the normal distribution.
3. Closed-Form c.d.f. – Unlike the normal distribution, the logistic distribution has an explicit cumulative distribution function, making it easier to work with in some applications.
4. Logistic Regression Link – The logistic function is used in classification problems to model probabilities.

11.6.4 Applications:

1. Logistic Regression – The c.d.f. is used as the activation function in binary classification models.
2. Economics and Growth Models – Models population growth and market saturation.
3. Survival Analysis and Reliability Engineering – Used to model failure times and life distributions.
4. Extreme Value Theory – Approximates distributions of maximum and minimum values.

11.7 PARETO DISTRIBUTION:

Let N be the number of persons whose income exceeds a given quantity x at a particular time. Then pareto has proposed that the number N can be assumed to follow a mathematical function given by

$$N = A \cdot x^{-a} -----------(1)$$

Where A and a (is known as pareto constant and shape parameter) are two constants. If the proportion of persons whose income exceeds x with a minimum income of k units is given by

$$p(x) = p(X \geqslant x) = \left(\frac{k}{x}\right)^a \text{--------------} (2)$$

$$\to \frac{k^a}{x^a} - k^a x^{-a} - Ax^{-a} (\because k^a - A).$$

$k > 0, a > 0, x \geqslant k.$

then the cumulative Distribution function is

$$F_X(x) = 1 - p(x) = 1 - \left(\frac{k}{x}\right)^a - - - - (3)$$
$$k > 0, a > 0; x \geqslant k.$$

the Equations (1) & (2) are almost similar in the form where in the place of A, we have k^a. Equation (3) satisfies the properties of a Distribution function. Its' density is given by

$$f(x) = \frac{a k^a}{x^{a+1}} (a > 0, x \geqslant k > 0) - - - - - (4)$$

$= 0$ otherwise

(4) is called the P.d. f of Pareto Distribution.

11.7.1 M.G.F of Pareto Distribution:

$$E(e^{tx}) = \int . e^{tx} \cdot f(x) dx$$
$$= \int_{-a}^{\infty} e^{tx} \cdot \frac{ak^a}{x^{a+1}} dx$$

$$ak^a \int_k^{\infty} e^{tx} \cdot x^{-a-1} dx$$
$= \text{put } x - k = z; dx = dz$
$$= ax^a \int_0^{\infty} e^{tx+tz} (z + k)^{-a-1} dz$$
$$= ak^a \int_0^{\infty} e^{t(z+k)} (z + k)^{-a-1} dz$$

which we can't write in closed form. rth moment is given by

$$E(x^r) = \int_k^{\infty} x^r \cdot f(x) dx$$
$$= \int_k^{\infty} x^r \cdot \frac{ak^a}{x^{a+1}} dx$$
$$= ak^a \int_k^{\infty} x^r \cdot x^{-a-1} dx$$
$$= ak^a \int_k^{\infty} x^{r-a-1} \, dx \left(\because \int x^x = \frac{x^{x+1}}{n+1}\right)$$

$x^{r-a} =$ finit moments does not exists

If r $a \geq 0$

$\frac{1}{(\infty)^{a-r}} = 0$ if $r - a < 0 \Rightarrow r < a$.

if $r < a$

$$\frac{ak^r}{r-a}(.\,k^{r-a}) = (-1)^{r-a}\frac{ax^a}{r-a}x^{r-a}$$
$$= \frac{ak^r(1)^{r-a}}{r-a} = \frac{ax^n}{a-1}$$
$$\text{Hean } E(x) = \frac{ak}{a-1} \cdot = \mu_1' \text{ if } a > 1$$
$$F(x^2) = \frac{ak^2}{a-2} = \mu_2', \text{ if } a > 2$$
$$\therefore\ u_2 = v(x) = E(x^2) - [\mathrm{E}(x)]^2.$$
$$= \frac{ak^2}{a-2} - \frac{a^2k^2}{(a-1)^2}$$
$$= ak^2\left\{\frac{1}{a-2} - \frac{a}{(a-1)^2}\right\}$$
$$= ak^2\left[\frac{(a-1)^2 - a(a-2)}{(a-1)^2(a-2)}\right]$$
$$= ak^2\left\{\frac{a^2 - 2a + 1 - \alpha^2 + 2\alpha}{(a-1)^2(a-2)}\right\}$$
$$= \frac{ak^2}{(a-1)^2(a-2)}$$

For the other moments.

$$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2\mu_1^3$$
$$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1' - 3\mu_1^{14}$$
$$\beta_1 = \frac{\mu_3^2}{\mu_2^3};\ \beta_2 = \frac{\mu_4}{\mu_2^2}.$$

11.7.2 Mean Deviation about Mean:

$$\mathrm{E}|x - \mathrm{E}(x)| = \int_k \left|x - \frac{ak}{a-1}\right| f(x)dx$$
$$= \int_{\frac{ak}{a-1}}^{\infty} \left(\mathrm{x} - \frac{ak}{a-1}\right) f(x)dx + \int_k^{a^{-1}} \left(\frac{ak}{a-1} - x\right) f(x)dx$$
$$x > \frac{ak}{a-1}$$
$$\frac{\int_{ak}^{a-1}}{\mathrm{a}-1} xf(x)dx - \int_{\frac{ak}{a-1}}^{\infty} \frac{ak}{a-1} f(x)dx + \int_k^{\frac{ak}{a-1}} \frac{ak}{a-1} f(x)$$
$$-\int_k^{a-1} xf(x)dx$$
$$-\ 2k(a-1)^{-1}(1 - a^{-1})^{a-1} \text{ and}$$

$$\frac{\text{Mean Deviation}}{\text{Standard Deviation}} = 2(1 - 2a^{-1})^{\frac{1}{2}}(1 - a^{-1})^{a-1}$$

$$\frac{\text{Mean Deviation}}{\text{Standard Deviation}} = 2(1 - 2a^{-1})^{\frac{1}{2}}(1 - a^{-1})^{a-1} \qquad (a>2)$$

The value of this ratio is 0.513 when $a = 3$.
When $a = 4$, this value is 0.597
As a tends to infinite, the ratio tend to $2e^{-1} = 0.736$

$$\text{M. G. F} \sum_{r=0}^{\infty} u_r' \frac{t^r}{r!}$$
$$= \sum_{x=0}^{a} x_{u.x^n}^{a-x} \cdot \frac{t^{n'}}{n!}$$
$$= a \sum_{x=0}^{\infty} \frac{(kt)^{n!}}{n!\,(a-x)}$$

which is not in continued form. Additive property is not satisfied.

11.8 CONCLUSION:

Understanding the Laplace, Weibull, Logistic, and Pareto distributions provides valuable insights into real-world data modeling. These distributions help in reliability analysis, risk assessment, machine learning, and economics. Mastering their properties enables effective statistical modeling, allowing for better decision-making in diverse applications such as engineering, finance, healthcare, and social sciences.

11.9 SELF ASSESSMENT QUESTIONS

1. What are the key characteristics of the Laplace, Weibull, Logistic, and Pareto distributions, and how do they differ from each other?
2. How does the shape and behavior of the Weibull distribution change based on its parameter values, and why is it widely used in reliability analysis?
3. In what scenarios is the Logistic distribution preferred over the Normal distribution, and what are its key advantages?
4. Why is the Pareto distribution considered a heavy-tailed distribution, and how is it applied in modeling wealth and risk analysis?
5. What are the key real-world applications of each of these distributions, and how do their properties make them suitable for different fields?

11.10 SUGGESTED BOOKS FOR FURTHER READING:

1. "Continuous Univariate Distributions, Volume 1" – Norman L. Johnson, Samuel Kotz, N. Balakrishnan (Wiley)
2. "Statistical Distributions" – Merran Evans, Nicholas Hastings, Brian Peacock (Wiley)
3. "A First Course in Probability" – Sheldon Ross (Pearson)
4. "Mathematical Statistics with Applications" – William Mendenhall, Robert J. Beaver, Barbara M. Beaver (Cengage Learning)
5. "An Introduction to Probability Theory and Its Applications" – William Feller (Wiley)

Dr. I.B.N. Hima Bindhu

LESSON -12

ORDER STATISTICS AND THEIR PROBABILISTIC PROPERTIES

OBJECTIVES

After completing this lesson, students will be able to

- Understand the Concept of Order Statistics
- Learn what order statistics are and why they are essential.
- Understand how they represent the sorted values from a random sample.
- Identify the first order statistic (minimum), last order statistic (maximum), and intermediate order statistic within a given sample.
- Derive the distribution function of single order statistic
- Analyze joint density functions

STRUCTURE

12.1 INTRODUCTION:

Order statistics play a crucial role in probability theory and statistical inference. They refer to the values obtained by arranging a random sample in increasing order. These statistics provide valuable insights into the extreme values, median, range, and variability of a dataset, making them useful in various real-world applications such as reliability analysis, risk assessment, and quality control.

12.2 ORDER STATISTICS:

Definition:

Let $(X_1, X_2, \dots X_n)$ be n-dimensional random vector and $(x_1, x_2, \dots x_n)$ be an n-tuple assumed by $(X_1, X_2, \dots X_n)$. Now, by arranging $x_1, x_2, \cdots; x_n$ in increasing order of magnitude, we get

$$X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$$

where $X_{(1)} = \min(x_1, x_2, \dots, x_n)$, i.e., $X_{(1)}$ is the smallest value in $x_1, x_2, \dots\dots x_n; X_{(2)}$ is the second smallest value of $x_1, x_2, \dots \cdots x_n$ and so on $X_{(n)} = \max(x_1, x_2, \dots x_n) \cdot X_{(i)}$ s $(i = 1,2, \dots n)$ are dependent because of the inequality relations among them.

Definition : The function $X_{(r)}$ of $(X_1, X_2, \dots X_n)$ that takes on the value x_r in each possible sequence $(x_1, x_2, \dots, x_n)$ of values assumed by $(X_1, X_2, \dots X_n)$is known as the rth order statistic or the statistic of order ' r '. $\{X_{(1)}, X_{(2)}, \cdots, X_{(n)}\}$ is called the set of order statistics for $(X_1, X_2, \dots X_n)$

Note: Let $(X_1, X_2, \dots X_n)$be an n- dimensional random variable. Let $X(r), 1 \leq r \leq n$ be the order statistic of order ' r '. Then $X_{(r)}$ is also random variable.

12.3 DISTRIBUTION FUNCTION OF A SINGLE ORDER STATISTIC:

Let $X_1, X_2, \dots X_n$ be n independent and identically distributed variates each with distribution function $F(x)$ and $\{X_{(1)}, X_{(2)}, \cdots, X_{(n)}\}$ be set of order statistics for $(X_1, X_2, \dots X_n)$. If $F_r(x), \mathrm{r} = 1,2, \dots \dots n$ denote the cumulative distribution function $(c \cdot d \cdot f.)$ of the r^{th}-order statistic $X_{(r)}$.

12.3.1 The c.d.f of the smallest order statistic $X_{(1)}$ is

$$F_1(x) = P[X_{(1)} \leq x]$$
$$= 1 - p[X_{(1)} > x]$$
$$= 1 - p[x_i > x, \forall i]$$
$$[\because X_1 \leq X_{(2)} \leq \cdots \leq X_{(n)}]$$

$$= 1 - \prod_{i=1}^{n} p(x_i > x)[\because X_i' s\ i = 1,2, \dots n \text{ an indenpendent}]$$

$$= 1 - \prod_{i=1}^{n}(1 - P(x_i \leq x)$$

$$=1- [1-F(x)]^n \quad [\because X_i' s\ i = 1,2, \dots n \text{ are i.i.d r.v's}]$$

Hence, the c. d. f of smallest order statistic $X_{(1)}$ is

$$X_{(1)} = 1 - [1 - F(x)]^n \quad \dots\dots\dots\dots\dots \quad (1)$$

12.3.2 The c. d. f of largest order statistics $X_{(n)}$ is

$$
\begin{aligned}
F_n(x) &= P[X_{(n)} \leq x] \\
&= P[X_i \leq x, i = 1,2, \ldots \ldots \ldots, n] \\
&= P(X_1 \leq x \cap X_2 \leq x \cap X_3 \leq x \cdots \cap X_n \leq x) \\
&= P(X_1 \leq x) \cdot P(X_2 \leq x) \ldots P(X_n \leq x) \\
&= \prod_{i-1}^{n} P(X_i \leq x) [\because X_i \text{ s an independent }] \\
&= \prod_{i=1}^{n} F(x_i) = [F(x)]^n [\because X_i \text{ is } i \cdot i \cdot d \cdot r \cdot \text{v's}]
\end{aligned}
$$

Hence the c. d. f of largest order statistics $X_{(n)}$ is $[F(x)]^n$ ……………….. (2)

12.3.3 The $c \cdot d \cdot f$ of r^{th} order statistic is

$$
\begin{aligned}
F_r(x) &= P(X_{(r)} \leq x) \\
&= P[\text{ at least } r \text{ of the } X_i \text{ 's are } \leq x] \\
&= \sum_{j=r}^{n} P[\text{ exactly } j \text{ of the } n, X_i \text{ 's are } \leq x]
\end{aligned}
$$

$$\sum_{j=r}^{n} \binom{n}{j} [F(x)]^j (1 - F(x)]^{n-j} \quad [\text{ by using binomial probability model }]$$

$\therefore c \cdot d \cdot f \cdot$ of r^{th} order statistic: is $\sum_{j=r}^{n} \binom{n}{j} [F(x)]^j (1 - F(x)]^{n-j}$ ……………………… (3)

The $c \cdot d \cdot f$ of r^{th} order statistic can be written as

$$F_r(x) = I_{F(x)}(r, n - r + 1) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.. \quad (4)$$

Where $I_y(a, b) = \int_0^y \frac{1}{\beta(a,b)} t^{a-1}(1-t)^{b-1} dt = \frac{1}{\beta(a,b)} \int_0^t t^{a-1}(1-t)^{b-1} dt$ …………… (5)

is the "incomplete beta function" and values if this function have been tabulated in Biometrika tables of pearson & Hartley. Hence the probability points of an order statistic can be obtained with the help of incomplete beta function.
Note: Taking $r = 1$ in (3), we get the $c \cdot d \cdot f$ of smallest order statistic

$$
\begin{aligned}
F_1(x) &= \sum_{j=1}^{n} \binom{n}{j} [F(x)]^j [1 - F(x)]^{n-j} \\
&= \sum_{j=0}^{n} \binom{n}{j} [F(x)]^j [1 - F(x)]^{n-j} - \sum_{j=0}^{J} \binom{n}{j} [F(x)]^j [1 - F(x)]^{n-j} \\
&= 1 - [1 - F(x)]^n \; [\because \text{ Total Prob is } 1]
\end{aligned}
$$

Similarly by taking $r = n$, we get the $c.d.f$ of largest order statistic is

$F_n(x) = [F(x)]^n$

12.4 PROBABILITY DENSITY FUNCTION (P. D. F.) OF r^{th} ORDER STATISTIC:

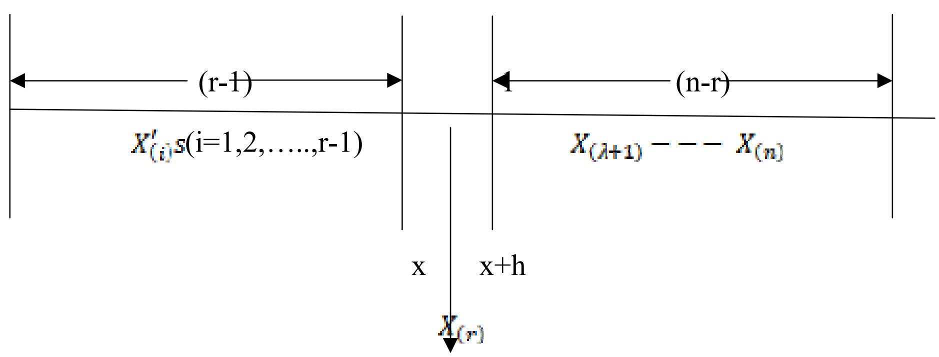
The $p \cdot d \cdot f.$ of a random variable x can be defined as

$$f(x) = \lim_{h \to 0} \left[\frac{F(x+h) - F(x)}{h}\right] = \lim_{h \to 0} \left(\frac{p(x < x \le x+h)}{h}\right] = \frac{dF(x)}{dx} = F'(x) \ldots \ldots (1)$$

Using this concept, the $p \cdot d \cdot f.$ of r^{th} order Statistic is

$$f_r(x) = \lim_{h \to 0} \left\{\frac{p(x \le x(r) \le x+h}{h}\right\} \quad -------- \quad (2)$$

Here the event $E: x < x_{(r)} \le x + \text{h}$ can be expressed as:



From the above figure

$X_i \le x$ for $(r-1)$ of the $X_{(i)}$'s

and $x < X_i \le x + h$ for one $X_{(i)}$.
and $X_i \geqslant x + h$ for the remaining $(n-r)$ of the $X_{(i)}$'s.
Using the multinomial probability law, we have

$$P(x \leqslant x_r \le x+h) = \frac{n!}{(\text{r}-1)!\, 1!\, (n-\text{r})!} p_1^{r-1} \cdot p_2^1 \cdot p_3^{n-r} \qquad (3)$$

Where $\qquad p_1 = P(x_i \le x) = F(x)$

$P_2 = P(x < X_i \le x+h) = F(x+h) - F(x)$
and $P_3 = P(X_i \geqslant x+h) = 1 - P(\text{X}_i \le x+h) = 1 - F(x+h)$

Substituting these in (3), we get

$$\text{f}(x) = \lim_{h \to 0} \left\{\frac{\{p(x < X_{(r)} \le x+h)}{h}\right\} = \frac{n!}{(\text{r}-1)!\, 1!\, (n-\text{r})!} [F(x)]^{r-1} \lim_{h \to 0} \left[\frac{F(x+h) - F(x)}{h}\right]^1 \lim_{h \to 0} \left[\frac{1 - F(x+h)}{h}\right.$$

$$= \frac{n!}{(r-1)!\,(n-r)!} [F(x)]^{r-1} \cdot f(x) \cdot [1 - F(x+h)]^{n-r}$$

12.5 JOINT $P \cdot d \cdot f \cdot$ OF ORDER STATISTICS:

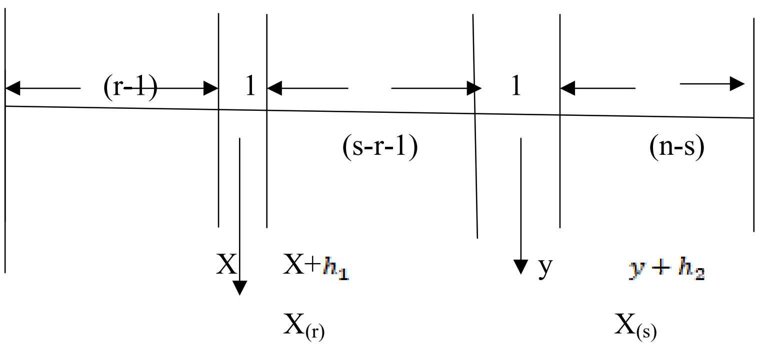
The joint probability density function of order statistics helps analyze the relationships between multiple ranked values in a sample.

12.5.1 Joint $P \cdot d \cdot f \cdot$ of two order statistics:

Let $f_{r,s}(x,y)$ be the p.d.f. of two order statistics $X_{(r)}$ and $X_{(s)}$ $(1 \le r < s \le n)$ of $\{X_{(1)}, X_{(2)}, \dots, X_{(n)}\}$, then

$$f_{rs}(x,y) = \lim_{\substack{h_1 \to 0 \\ h_2 \to 0}} \left\{ \frac{p[x \le X_{(r)} \le x + h_1 \cap y < X_{(s)} \le y + h_2]}{h_1 . h_2} \right\} \qquad (1)$$

Here the event $E: \{x < x_{(r)} \le x + h_1 \cap y < X_{(s)} \le y + h_2\}$ can be expressed as



From the figure, we have $X_i \le x$ for $(r-1)$ of the $x_{(i)}'s$

$x \le X_i \le x + h_1$ for one $X_{(i)}$,

$x + h_1 \le X_i \le y$ for $(s - r - 1)$ of $X_{(i)}'s$,

$y \le X_i \le y + h_2$ for one $x_{(i)}$

and $x_i > y + h_2$ for $(n - s)$ of the $X_i(i)'s$

By using multinomial probability, we have.

$$P\{x < X(r) \le x + h_1 \cap y < X(s) \le y + h_2\} = \frac{n!}{(r-1)!\,1!\,(s-r-1)!\,1!\,(n-s)!} p_1^{r-1} . p_2' \cdot p_3^{s-r-1} \cdot p_4' \cdot p_5^{n-1} . (2$$

$p_1 = P(X_i \le x) = F(x)$, $p_2 = P(x < X_i \le x + h_1) = F(x + h_1) - F(x)$

$p_3 = P(x + h_1 < X_i \le y) = F(y) - F(x + h_1)$

$p_4 = P(y \le X_i \le y + h_2) = F(y + h_2) - F(y)$

$P_5 = P(X_i > y + h_2) = 1 - P(X_i \le y + h_2) = 1 - F(y + h_2)$

By. Substituting (2) in (1), we get

$$f_{r_s}(x,y) = \lim_{h_1\to 0, h_2\to 0}\left\{\frac{P(E)}{h_1 h_2}\right\}.$$

$$= \frac{n!}{(r-1)!\cdot(s-r-1)!\,(n-s)!}[F(x))^{r-1}\cdot \lim_{h_1\to 0}\left\{\frac{F(x+h_1)-F(x)}{h_1}\right\}'\cdot \lim_{h_1\to 0}\left[\frac{F(y)-F(x+h_1)}{h_1}\right]^{s-r-1}$$

$$\lim_{h_2\to 0}\left\{\frac{F(y+h_2)-F(y)}{h_2}\right\}'\cdot \lim_{h_2\to 0}\{1-F(y+h_2)]^{n-s}$$

$$= \frac{n!}{(r-1)!\cdot(s-r-1)!\,(n-s)!}[F(x))^{r-1}\cdot f(x)\,[F(y)-F(x)]^{s-r-1}f(y).\,[1-F(y)]^{n-s} \quad (3)$$

Hence the density given in (3) is the $P\cdot d\cdot f$ of two order statistics $X_{(r)}$ and $X_{(s)}$

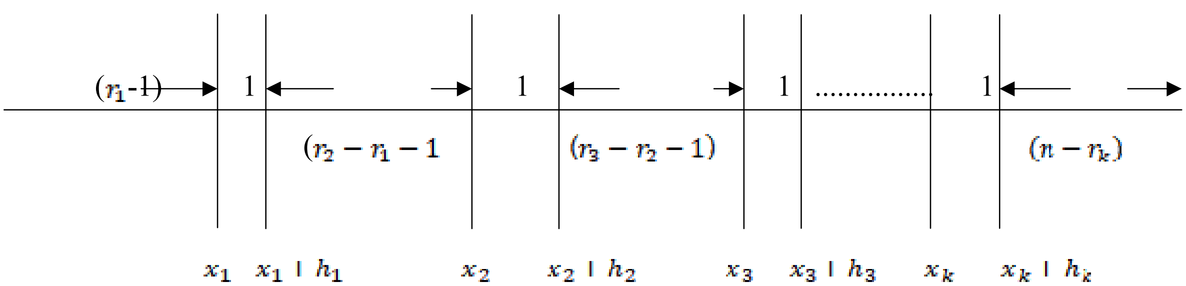
$(1 \le r < s \le n)$ of the order statistics $(X_{(1)}, X_{(2)}, \dots X_{(n)})$.

12.5.2 Joint P. d. f. of K-order statistics:

We can extend the result of joint P. d. f of two order statistics (using multi normal Probabilities) to get the $p.d.$ f of k-order statistics $X_{(r_1)}, X_{(r_2)}, \cdots, X_{(r_k)}$ when $1 \le r_1 < r_2 < \cdots < r_k \le n$ and $1 \le k \le n$ is for $x_1 \le x_2 \le \cdots < x_k$ as

$$f_{r_1,r_2,\cdots r_k}(x_1, x_2, \cdots x_k) = \frac{n!}{(r_1-1)!\,(r_2-r1-1)!\,(r_k-r_{k-1}-1)!\,(n-r_k)!}(F(x_1)]^{r_1-1}\cdot f(x_1).$$

$$[F(x_2)-F(x_1)]^{r_2-r_1-1}\cdot f(x_2)\cdots \dots f(x_k)\cdot[1-F(x_k)]^{n-r_k}$$



12.5.3 Joint p.d.f of all the n-order statistics:

Theorem: The joint $p\cdot d\cdot f$ of $\{x_{(1)}, x_{(2)}, \dots, x_{(n)}\}$ is given by

$$f(x_{(1)}, x_{(2)}, \dots x_{(n)}) = n!\,g(x_{(r)})\big), \text{if}(x_{(1)} < x_{(2)} < \cdots < x_{(n)}).$$
$$= 0 \text{ otherwise}.$$

Proof:- Let as assume that $X_1, X_2, \dots X_n$ be i.i.d r. v.'s of the continuous type with p. d. f $g(x)$ and $\{X_{(1)}, X_{(2)}, \cdots X_{(n)})\}$ be the set of order statistics for $X_1, X, \dots, X_n$. Since the X_i are all continuous type r. v's, it follows with probability 1 that $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$, if $(x_1, x_2, \dots x_n)$ be set of values assumed by $(X_1, X_2 \dots X_n)$.

The transformation from $(X_1, X_2, \dots, X_n)$ to $(X_{(1)}, X_{(2)}, \dots, X_{(n)})$ is not one - to -one, since a total of $n!$ possible arrangements of $x_1, x_2, \dots x_n$ in increasing order of magnitude. Thus there are $n!$ inverses to the transformation. For example, one of the n ! permutations might be

$$x_2 < x_1 < x_5 < x_{n-1} < x_3 < -1 \cdots < x_n < x_4.$$

Then the corresponding inverse is

$$x_2 = x_{(1)}, x_1 = x_{(2)}, x_5 = x_{(3)}, x_{n-1} = x_{(4)}, x_3 = x_{(5)}, \dots x_n = x_{(n-1)}, x_4 = x_{()}$$

The Jacobian of the transformation is the determinant of an $n \times n$ identity matrix with rows rearranged, since each $x_{(i)}$ equals one and only one of $X_1, X_2, \dots, X_n$. Therefore $J = \pm$ and hence $f\left(x_{(2)}, x_{(1)}, x_{(5)}, \cdots x_{(n-1)}\right) = |J| . \prod_{i=1}^{n} g(x_i) \quad , x_{(1)} < x_{(2)} < \cdots < x_{(n)}$

The same expression holds to each of the n! Arrangements.

$$\begin{aligned} \therefore f\left(x_{(1)}, x_{(2)}, \dots \cdot x_{(n)}\right) &= n! \prod_{i=1}^{n} g(x_i) \\ &= n!\, g(x_1) g(x_2) \cdots g(x_n), \text{ if } x_{(1)} < x_{(2)} \dots < x_{(n)} \\ &= 0, \text{ otherwise} \end{aligned}$$

12.6 SOLVED EXAMPLES:

Example (1) : Let $X_1, X_2, \dots X_n$ be i.i.d r. v's with common p d f

$$\begin{aligned} g(x) &= 1, \text{ if } 0 < x < 1 \\ &= 0, \text{ otherwise} \end{aligned}$$

then the Joint p d f of the order statistics $X_{(1)}, X_{(2)}, \cdots X_{(n)}$ is

$$\begin{aligned} f\left(x_{(1)}, x_{(2)}, \cdots x_{(n)}\right) &= n! \prod_{i=1}^{n} g(x_i) \\ &= n!\, g(x_1) \cdot g(x_2) \cdots g(x_n) \\ &= n! [g(x_i)]^n \cdots [\because X_i \text{ are i.i.d}] \\ &= n! (1)^n [\because x_i \text{'s are i.i.d }] \\ &= n! \\ \therefore f\left(x_{(1)}, x_{(2)}, \cdots, x_{(n)}\right) &= n!, 0 < x_{(1)} \leq x_{(2)} < \cdots < x_{(n)} < 1 \\ &= 0 \text{ otherwise.} \end{aligned}$$

Example (2): - If X_1, X_2, X_3, X_4 are continuous i.i.d $r \cdot v$'s with $pdf\, f(x)$ and the joint pdf of $X_{(1)}, X_{(2)}, X_{(3)}, X_{(4)}$ is

$$\begin{aligned} f\left(x_{(1)}, x_{(2)}, x_{(3)}, x_{(4)}\right) &= 4! \cdot f(x_1) \cdot f(x_2) \cdot f(x_3) \cdot f(x_4), \text{if } x_{(1)}, < x_{(2)}, < x_{(3)} < x_{(4)} \\ &= 0, \text{otherwise} \end{aligned}$$

then find the marginal $p \cdot d \cdot f \cdot$ of $X_{(2)}$.
Sol: The marginal $p \cdot d \cdot f \cdot of \cdot X_{(2)}$ is

$$f_2(x) = 4! \iiint g(x_1) g(x_2) \cdot g(x_3) g(x_4) \; dx_1 \, dx_3 \, dx_4$$

$$f_2(x) = 4!\, g(x_2) \int_{-\infty}^{x_2} \int_{x_2}^{x_3} [\int_{x_3}^{\infty} g(x_4)dx_4]\, g(x_3) g(x_1) dx_3\, dx_1$$

$$= 4!\, g(x_2) \int_{-\infty}^{x_2} \{ \int_{x_2}^{\infty} [1 - G(x_3)] \cdot g(x_3) dx_3 \Big\} g(x_1) dx_1$$

$$= 4!\, g(x_2) \int_{-\infty}^{x_2} \frac{[1 \quad G(x_2)]^2}{2} g(x_1) dx_1 \qquad = 4! \frac{[1 - G(x_2)]^2}{2!} \cdot G(x_2) \cdot g(x_2) . x_{(1)} < x_{(2)} < x_{(3}$$

$$= 4!\, g(x_2) \frac{[1 - G(x_2)]^2}{2!} G(x_2)$$

$$\therefore f_2(x) = f(x_{(2)})$$

$$< x_{(4)}.$$

Example (3):- Let $X_1, X_2, \ldots X_n$ be $i.i.d$ random variables with common p d f

$g(x) = 1$, if $0 < x < 1$
$= 0$, otherwise

then the $p \cdot d \cdot f$ of r th order statistic of $X, X_{(2)}, \ldots X_{(n)}$, is

$$f(x_r) = \frac{n!}{(r-1)!\,(n-r)!} x_r^{r-1} (1 - x_r)^{n-r}, 0 < x_r < 1 \; (1 \le r \le n)$$
$$= 0, \text{ otherwise.}$$

The joint $p \cdot d \cdot f$ of $x_{(r)}$ and $x_{(s)}$ is $(1 \le r < s \le n)$

$$f(x_{(r)}, x_{(s)}) = \frac{n!}{(r-1)!\,(s-r-1)!\,(n-s)} x_r^{r-1} (x_r - x_s)^{s-r-1} (1 - x_k)^{n-s} \; ; x_r < x_s$$
$$= 0. \text{ otherwise.}$$

12.7 CONCLUSION:

Order statistics focus on arranging sample values in increasing order and studying their properties. The smallest value in a sample is the first order statistic, while the largest is the last order statistic. Any value in between represents a general order statistic. The distribution function of an order statistic provides the probability that it does not exceed a given value, which helps in estimating extreme values and medians. The probability density function describes the likelihood of an order statistic taking a specific value. The joint probability density function examines the probability of multiple order statistics occurring together. These concepts are widely applied in reliability testing, quality control, extreme value analysis, and statistical inference.

12.8 SELF ASSESSMENT QUESTIONS:

1. What are order statistics, and how are they defined in a given sample?
2. How does the distribution function of an order statistic help in probability analysis?
3. What is the significance of the probability density function of a single order statistic?
4. How is the probability density function of a single order statistic derived from a given sample?
5. What does the joint probability density function of order statistics represent?

6. How can order statistics be used to analyze extreme values in a dataset?
7. In what ways are order statistics applied in reliability testing and quality control?
8. How does the joint probability density function of order statistics differ from the probability density function of a single order statistic?
9. Why are order statistics important in survival analysis and risk assessment?
10. How can order statistics be used in real-world scenarios such as finance, engineering, and medical research?

12.9 SUGGESTED BOOKS:

1. "Order Statistics" – H. A. David and H. N. Nagaraja
2. "Introduction to Probability and Statistics" – J. S. Milton and J. C. Arnold
3. "Mathematical Statistics with Applications" – W. Mendenhall, R. J. Beaver, and B. M. Beaver
4. "An Introduction to Probability Theory and Its Applications" – William Feller
5. "Probability and Statistical Inference" – Robert V. Hogg, Elliot A. Tanis, and Dale L. Zimmerman
6. "Mathematical Statistics" – John E. Freund

Dr. I.B.N. Hima Bindu

LESSON - 13

DISTRIBUTION OF RANGE AND ITS APPLICATIONS

OBJECTIVES:

After completing this lesson, students will be able to

- Understand the definition of range in the context of order statistics.
- Deriving the distribution of range
- Apply range of Order Statistics to rectangular and uniform distributions.

STRUCTURE:

13.1 INTRODUCTION:

The range of a dataset is a measure of how spread out the values are. It is calculated as the difference between the largest and smallest values in a sample. For example, if we record the daily temperatures in a city for a week and the highest temperature is 35°C while the lowest is 20°C, the range would be 15°C. The distribution of the range refers to the probability of different possible values of this range when data is collected from a particular probability distribution. It helps in understanding the expected spread of data in repeated samples.

13.2 DISTRIBUTION OF RANGE:

Let us consider the joint p.d.t. of the order statistics $x_{(r)} = x$ (say) and $x_{(s)} = y (r < s)$ of $X_{(1)}, X_{(2)} \cdots X_{(n)}$.

$$f(x, y) = \frac{n!}{(r-1)! \cdot (s-r-1)!\,(n-s)!} [F(x)]^{r-1} \cdot f(x) [F(y) - F(x)]^{s-r-1} f(y). F[1 - F(y)]^{n-s} \quad (1)$$

Let $\omega_{rs} = X_{(s)} - X_{(r)}, (r < s)$ be the difference of two order statistics. To obtain the distrain of w_{rs} , consider the joint $p \cdot d \cdot f$. of $x_{(r)}$, ant $x_{(s)}$ given (1) and transform $[X_{(r)}, X_{(s)}]$ i.e., $[x, y]$ to the new variables ω_{rs} and $X_{(r)}$ such that

$w_{rs} = y - x, x_{(r)} = x \Rightarrow y = x + w_{rs}$ and $x = x$

∴ Jacobian transformation is

$$J = \frac{\partial(x, y)}{\partial(x, w_{rs})} = \begin{vmatrix} 1 & 1 \\ 0 & 1 \end{vmatrix} = 1 \Rightarrow |J| = 1$$

∴ The joint $p \cdot d$.f .of x, ω_{rs} is

$$g(x, w_{rs}) = |J| \cdot f(x, y).$$
$$= c_{rs} \cdot [F(x)]^{r-1} \cdot f(x) \cdot [F(x + w_{rs}) - F(x)]^{s-r-1} f(x + w_{rs}).[1 - F(x + w_{rs})]^{n-s} \qquad (2)$$

where $x=x, y=x + w_{rs}, \; c_{rs} = \frac{n!}{(r-1)!\cdot(s-r-1))(n-r)!}$

Integrating $g(x, w_{rs})$ w.r. to x over $-\infty$ to ∞, we get

$$g(w_{rs}) = c_{rs} \int_{-\infty}^{\infty} \{[F(x)]^{r-1} f(x) \cdot [F(x + w_{rs}) - F(x)]^{s-r-1} \cdot f(x + w_{rs}) \cdot [1 - F(x + w_{rs})]^{n-s}\} dx$$

$$\text{where } c_{rs} = \frac{n!}{(r-1)! \cdot (s - r - 1))(n - r)!} \qquad (3)$$

By taking $r = 1$ and $s = n$ in the above result (3), we get the p.d.f of the range $\omega = X(n)$ - $X_{(1)}$ which

$$g(\omega) = \frac{n! n(n-1)}{0! (n-1-1)! \, 0!} \int_{-\infty}^{\infty} f(x) \cdot [F(x + \omega) - F(x)]^{n-2} \cdot f(x + \omega) dx \; ; \omega \geq 0 \qquad (4)$$

Now the distribution function of the range (from (4)) is

$$\varphi(\omega) = p(W \leq \omega) = \int_0^{\omega} g(u)du, \text{ where } g(u) \text{ is given in (4).}$$

$$= \int_0^{\omega} \left\{ n(n-1) \int_{-\infty}^{\infty} f(x)[F(x+u) - F(x)]^{n-2} f(x+u)dx \right\} du.$$

$$= n \int_{-\infty}^{\infty} f(x) \left\{ \int_0^{\omega} (x-1) \cdot f(x+u)[F(x+u) - F(x)]^{n-2} du \right\} dx.$$

$$= n \int_{-\infty}^{\infty} f(x) \left\{ \int_0^{\omega} d[F(x+u) - F(x)]^{n-1} \right\} dx$$

$$= n \int_{-\infty}^{\infty} f(x)[F(x+w) - F(x)]^{n-1} dx$$

$$\left\{ \begin{aligned} &since \frac{d}{du}(F(x+u) - F(x)]^{n-1} \\ &\quad = (n-1)[F(x+u) - F(x)]^{n-2} \cdot \frac{d}{du} \mathrm{F}(x+u) \\ &\quad = (n-1)[F(x+u) - F(x)]^{n-2} f(x+u) \end{aligned} \right.$$

∴ The distribution function of the Range is

$$\varphi(\omega) = n \int_{-\infty}^{\infty} f(x)[F(x+\mathrm{w}) - F(x)]^{n-1} dx, \omega \geqslant 0$$

and the p. d. f. of Range is.

$$g(\omega) = n(\mathrm{n}-1) \int_{-\infty}^{\infty} f(x)[F(x+\omega) - F(x)]^{n-2} f(x+\omega)dx, \omega \geq 0$$

13.3 SOLVED EXAMPLES:

Example 1: Suppose $x_1, x_2, \dots \dots \dots . x_n$ is sample of size n drawn from a distribution. Show that

$y_1 = \min(x_1, x_2, \dots \dots x_n)$ or the smallest order statistics of $x_1, x_2, \dots \dots \dots . x_n$ follows exponential distribution with parameter $n\theta$ if and only if each x_i follows exponential distain with parameter θ.

Sol: Necessary Condition:

Let $X_i (i = 1, 2, \cdots n)$ be $i \cdot i \cdot \mathrm{d}$ exponential variates with parameter θ and its density in given by

$$f(x) = \theta \cdot e^{-\theta x}, x \geqslant 0, \theta > 0 \qquad (1)$$

∴ The distribution function is

$$F(x) = P(\mathrm{X} \leq x) = \int_0^x f(z)dz = \int_0^x \theta . e^{-\theta z} dz = 1 - e^{-\theta x} \qquad (2)$$

Now the distribution function for of $y_1 - \min(x_1, x_2, \dots x_n)$ is given by

$$\varphi_{y1}(y) = P(y_1 \le y) = 1 - [1 - F(y)]^n : \left[\because y_1 \text{ is the smallest order statistic}\right.$$
$$= 1 \quad [1 \quad (1 \quad e^{-\theta y})]^n$$
$$= 1 - e^{n\theta y} \quad \ldots\ldots\ldots\ldots\ldots\ldots (3)$$

Hence (3) is the distribution function of exponential distribution with parameter $n\theta$, which implies y_1 (smallest order statistics) has exponential distribution with parameter $n\theta$.

Sufficient condition: Let $y_1 = \min(x_1, x_2, \ldots . x_n)$ *i.e.*, y_1 is the smallest order statistic follow exponential distribution with parameter $n\theta$, then we have to show that x_i, follows exponential distain with parameter θ,

$$P(Y_1 \le y) = 1 - e^{-n\theta y} \Rightarrow P(Y_1 > y) = e^{-n\theta y}.$$
$$\Rightarrow P[\min(x_1, x_2, \ldots, x_n) \geqslant y] = e^{\;n\theta y}$$
$$\Rightarrow P[x_i \geqslant y \cap x_2 > y \cap \ldots n x_n > y] = e^{-n\theta y}$$
$$\Rightarrow \prod_{i=1}^{n} P(x_i > y) = e^{-n\theta y} [\because x_i\text{'s are independent}]$$
$$\Rightarrow [P(x_i > y]^n = e^{-n\theta y} : [\because x_i\text{'s are i.i.}\theta]$$
$$\Rightarrow P(x_i \ge y) = e^{-\theta y}$$
$$\Rightarrow P(x_i < y) = 1 - e^{-\theta y}$$

which is the distribution function of the exponential distribution with parameter θ.

Hence the sufficient condition.

Example 2: Show that in odd samples of size n from Uniform distribution over $[0,1]$, the mean and variance of the distribution of median are $\frac{1}{2}$ and $\frac{1}{4(n+2)}$ respectively.
Ans :- The $p \cdot d \cdot f$ of uniform distribution over $[0,1]$ is.

$f(x) = 1, 0 \le x \le 1 \ldots\ldots\ldots\ldots\ldots (1)$

and the distribution function is
$F(x) = P(X \le x) = \int_0^x f(u)du = \int_0^x 1 \cdot du = x$ (2)
Let the sample size n$= 2m + 1$, where m is +ve integer $\geqslant 1$,
Then the sample be $x_{(1)}, x_{(2)}, \ldots x_{(m+1)}, \ldots . x_{(n)}$.
$\therefore$ The median observation is $X(m+1)$
Now consider the $p \cdot d \cdot f$ of r^{th} order statistic $x_{(r)}$ of the sample $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ is

$$f_r(x) = \frac{1}{\beta(r, n-r-1)} [F(x))^{r-1} f_{(x)} [1 - F(x)]^{n-r} \qquad (3)$$

$\therefore$ using (3), the $p.d.f$ of the $X_{(m+1)}$ (median) of $X_{(1)}, X_{(2)}, \ldots . X_{(m+1)}, \ldots, X_{(2m+1)}$ is mean is

$$f_{m+1}(x) = \frac{1}{\beta(m+1,m+1)}\, x^m \cdot (1-x)^m$$

$$E[X_{(m+1)}] = \int_0^1 x \cdot \frac{1}{\beta(m+1,m+1)}\, x^m \cdot (1-x)^m dx.$$

$$\begin{bmatrix} \because r = m+1 \\ n - \mathrm{r} = 2m+1-(m+1) = m \\ n-r+1-m+1 \end{bmatrix}$$

$$= \frac{1}{\beta(m+1,m+1)}\int_0^1 x^{m+2-1} \cdot (1-x)^{m+1-1} dx$$

$$= \frac{\beta(m+2,m+1)}{\beta(m+1,m+1)} = \frac{\Gamma(m+2)\cdot\Gamma(m+1)}{\Gamma(m+3)} \cdot \frac{\Gamma(2m+2)}{\Gamma(m+1)\cdot\Gamma(m+1)}$$

$$\therefore \text{ mean } = \frac{m+1}{2m+2} = \frac{1}{2} \text{ (cn simplification)}$$

similarly we get $E[X^2_{(m+1)}] = \frac{m+2}{2(2m+3)}$

$$\therefore \text{ variance of } X_{(m+1)} = E[X^2_{(m+1)}] - [E(x_{(m+1)})]^2$$

$$= \frac{m+2}{2(2m+3)} - \frac{1}{4} = \frac{1}{4(2m+3)} = \frac{1}{4(n+2)} \cdot \begin{bmatrix} n = 2m+1 \\ = 2m+1 \end{bmatrix}$$

13.4 APPLICATIONS OF DISTRIBUTION OF RANGE:

13.4.1 In Rectangular distribution:

The distribution of the range in a rectangular (uniform) distribution has several applications across different domains, particularly in statistics, reliability analysis, quality control, and extreme value theory. The following are some applications:

1. Quality Control and Engineering

- The range is often used as a measure of process variability in quality control.
- In statistical process control (SPC), the range of a sample is used in range charts (RR-charts) to monitor the consistency of production processes.
- When data follows a uniform distribution (e.g., sensor readings, certain manufacturing tolerances), the distribution of the range helps set control limits.

2. Extreme Value Analysis

- Since the range is the difference between the maximum and minimum values in a sample, studying its distribution is useful in extreme value theory.

- Applications include modeling environmental extremes (e.g., daily temperature range), financial risks (e.g., range of stock prices), and other cases where uniformity is assumed in the underlying process.

3. Reliability Engineering

- In reliability testing, if failure times are assumed to follow a uniform distribution over an interval, the distribution of the range helps in predicting time to failure spread.
- It is useful in stress testing to determine worst-case scenarios.

4. Signal Processing and Communications

- In signal processing, uniform noise models often arise, and the range distribution can be used to estimate signal variation and dynamic range.
- Applications include modulation techniques, image processing, and error detection.

5. Simulation and Random Sampling

- When simulating random variables from a uniform distribution, the expected range provides insights into sample variability.
- Used in Monte Carlo simulations for modeling random behavior in various applied fields.

6. Estimation and Statistics

- The range can serve as a simple estimator of variability in small samples when variance estimation is difficult.
- In some cases, the range is used as a statistic to estimate parameters of the uniform distribution itself (e.g., estimating the width of a uniform distribution).

13.4.2 In exponential distribution:

The distribution of the range in an exponential distribution has important applications in reliability analysis, survival analysis, queueing theory, and extreme value statistics. Since the exponential distribution is often used to model waiting times, failure times, and interarrival times, the range of a sample provides useful insights into variability, worst-case scenarios, and system reliability.

Key Applications of the Range in Exponential Distribution:

1. Reliability and Life Testing

- In reliability engineering, the exponential distribution models the time until failure of components in systems with a constant failure rate.
- The range (difference between the longest and shortest lifetimes in a sample) helps in estimating the variability of lifetimes.
- The distribution of the range can be used to:

  - Compare reliability between different components.
  - Set warranty periods based on observed failure ranges.

- Design stress tests to ensure product durability.

2. Extreme Value Analysis

- The range distribution in an exponential setting helps in modeling extremes of waiting times.
- Applications include:

- Predicting longest and shortest waiting times in a system (e.g., time between major earthquakes, longest vs. shortest service times in a queue).
- Analyzing rare events, such as extreme failure times in industrial machines.

3. Queueing Theory and Operations Research

- Many real-world systems involve queues where service times or interarrival times follow an exponential distribution (e.g., call centers, hospital emergency rooms).
- The range helps in:
- Estimating the maximum and minimum waiting times in a given period.
- Setting upper and lower bounds for queue management and resource allocation.
- Designing efficient scheduling systems to reduce waiting time variability.

4. Statistical Inference and Estimation

- The range of an exponential sample is sometimes used as an estimator of the scale parameter (λ\lambda), particularly in cases where traditional methods (like the mean) are unreliable.
- In parameter estimation, the distribution of the range can assist in:
- Estimating the rate parameter of an exponential process.
- Testing goodness-of-fit by comparing observed ranges with expected ranges.

5. Simulation and Monte Carlo Methods

- Many simulations involving stochastic processes use exponentially distributed waiting times.
- The range helps in analyzing variability in simulated processes.
- Example: Simulating hospital emergency room wait times—the range provides insight into how long the longest and shortest patients wait.

6. Signal Processing and Communication Systems

- In some communication systems, signal arrival times or packet transmission times follow an exponential distribution.
- The range of these times helps in evaluating network performance, such as:
- Delays in data transmission.
- Worst-case response times in real-time systems.

13.5 CONCLUSION:

The range of a dataset is the difference between the largest and smallest values. In statistics, understanding the distribution of the range helps in analyzing variability and extreme values. For a rectangular (uniform) distribution, the range follows a specific pattern based on order statistics. When values are taken from a uniform distribution, the range depends on the highest and lowest observations. This is useful in quality control and random number generation. In the exponential distribution, the range plays a key role in reliability studies and survival analysis. Since the exponential distribution models waiting times and lifetimes, analyzing the range helps in understanding failure rates and risk assessment. The study of range distribution is widely applied in inferential statistics, engineering, and decision-making processes.

13.6 SELF ASSESSMENT QUESTIONS:

1. What is the range of a dataset, and why is it important in statistical analysis?
2. How is the distribution of the range related to order statistics?
3. In a rectangular (uniform) distribution, how does the sample size affect the expected range?
4. What are some real-world applications of studying the range in a uniform distribution?
5. How is the range distributed in the case of an exponential distribution?
6. Why is the exponential distribution particularly useful in reliability and survival analysis?
7. How does the parameter of an exponential distribution influence the expected range?
8. Compare the behavior of the range in uniform and exponential distributions. How do they differ?
9. In which fields can the distribution of the range be applied, and why is it useful in those areas?
10. Can the distribution of the range be extended to other probability distributions? Give an example.

13.7 SUGGESTED BOOKS:

1. David, H. A., & Nagaraja, H. N. (2003). Order Statistics (3rd ed.). Wiley.
2. Arnold, B. C., Balakrishnan, N., & Nagaraja, H. N. (1992). A First Course in Order Statistics. SIAM.
3. Johnson, N. L., Kotz, S., & Balakrishnan, N. (1994). Continuous Univariate Distributions, Volume 1 (2nd ed.). Wiley.
4. Casella, G., & Berger, R. L. (2002). Statistical Inference (2nd ed.). Duxbury Press.
5. Mood, A. M., Graybill, F. A., & Boes, D. C. (1974). Introduction to the Theory of Statistics (3rd ed.). McGraw-Hill.

Dr. I.B.N.Hima Bindu