Unit-1

Probability

1.1

Algebra of Sets and Counting Methods

The algebra of sets and counting methods are useful in understanding the basic concepts of *probability*. These concepts are briefly reviewed from the point of view of probability.

Sets and Elements of sets: The fundamental concept in the study of the probability is the set.
A set is a well defined collection of objects and denoted by upper case English letters. The objects in a set are known as **elements** and denoted by lower case letters. A set can be written in two ways. Firstly, if the set has a finite number of elements, we may list the elements, separated by commas and enclosed in brackets. For example, a set A with elements $1, 2, 3, 4, 5$ and 6, it may be written as

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

Secondly, the set may be described by a statement or a rule. Then A may be written as

$$A = \{x \mid x \text{ is a natural number less than or equal to } 6\}$$

If x is an element of the set A, we write $x \in A$. If x is not a element of the set A, then we write $x \notin A$.

Equal Sets: Two sets A and B are said to be ***equal*** or ***identical*** if they have exactly the same elements and we write as $A = B$

Subset: If every element of the set A belong to the set B, $i.e.$, if $x \in A \implies x \in B$, then we say that A is a ***subset*** of B and we write $A \subseteq B$ (A is contained in B) or $B \supseteq A$ (B contians A). If $A \subseteq B$ and $B \subseteq A$, then $A = B$.

Null set: A ***null*** or an ***empty set*** is one which does not contain any element at all and denoted by ∅.

Note:

1. Every set is a subset it self
2. An empty set is a subset of every set.
3. A set containing only one elements is conceptually different from the element itself .
4. In all applications of set theory, especially in probability theory, we shall have a fixed set S (say), given in advance and we shall be concerned only with subsets of S. This set is referred to universal set.

1) **Union or sum:**

$$A \cup B = \{x |\ x \in A \ \ or \ \ x \in B\}$$

$$\bigcup_{i=1}^{n} A_i = \{x| \ \ x \in A_i \quad for\ at\ least\ one\ i = 1, 2, \dots, n\}$$

2) **Intersection or Product:**

$$A \cap B = \{x| \ \ x \in A \ \ and \ \ x \in B\}$$

$$\bigcap_{i=1}^{n} A_i = \{x|\ x \in A_i \ \ for\ all\ i = 1, 2, \dots, n\}$$

If $A \cap B = \emptyset$, then we say that A and B are **disjoint sets.**

3) **Relative Difference:** $A - B = \{x| \ \ x \in A \ \ and \ \ x \notin B\}$
4) **Complement:** $\bar{A} = S - A$

Algebra of Sets:

If A, B and C are subsets of a universal set S, then the following laws hold:

Commutative laws: $A \cup B = B \cup A,\ A \cap B = B \cap A$

Associative laws: $(A \cup B) \cup C = A \cup (B \cup C), (A \cap B) \cap C = A \cap (B \cap C)$

Distributive laws:

$A \cap (B \cup C) = (A \cap B) \cup (A \cap C), A \cup (B \cap C) = (A \cup B) \cap (A \cap C)$

Complementary laws: $A \cup \bar{A} = S, A \cap \bar{A} = \emptyset, A \cup S = S, A \cap S = A$

Difference laws: $A - B = A \cap \bar{B} = A - (A \cap B) = (A \cup B) - B,$

$A - (B - C) = (A - B) \cup (A - C), (A \cup B) - C = (A - C) \cup (B - C),$

$(A \cap B) \cup (A - B) = A, (A \cap B) \cap (A - B) = \emptyset$

De – Morgan's laws:

$\overline{A \cup B} = \bar{A} \cap \bar{B}, \overline{A \cap B} = \bar{A} \cup \bar{B}$

$\overline{\bigcup_{i=1}^{n} A_i} = \bigcap_{i=1}^{n} \bar{A}_i$ and $\overline{\bigcap_{i=1}^{n} A_i} = \bigcup_{i=1}^{n} \bar{A}_i$

Involution law: $\overline{(\bar{A})} = A$

Idempotent law: $A \cup A = A, A \cap A = A$

Class of Sets: A group of sets will be termed as a class of sets. We shall define some useful types of classes used in probability.

Field: A field $\mathbb{F}$ (or algebra) is a non – empty class of sets which is closed under the formation of finite unions and under complementation. Thus,

(i) $A \in \mathbb{F},\ B \in \mathbb{F} \implies A \cup B \in \mathbb{F}$ and
(ii) $A \in \mathbb{F} \implies \bar{A} \in \mathbb{F}$

***σ* – Field:** A σ – field $\mathbb{B}$ (or σ – algebra) is a non – empty class of sets that is closed under the formation of countable union and complementation. Thus,

(i) $A_i \in \mathbb{B},\ i = 1,2,\ldots, \implies \bigcup_{i=1}^{n} A_i \in \mathbb{B}$

(ii) $A \in \mathbb{B} \implies \bar{A} = \mathbb{B}$

Fundamental Principle of Addition (Principle of inclusion- exclusion)

Let $A_1, A_2, \dots A_m$ be m sets and the elements in each sets are different. Then the number of ways of selecting an element from A_1 or A_2 or …A_m is given by

$$n\left(\bigcup_{i=1}^{m} A_i\right) = \sum_{i=1}^{m} n(A_i) - \sum_{\substack{i=1 \\ i<j}}^{m} \sum_{j=1}^{m} n\left(A_i \cap A_j\right) + \sum_{\substack{i=1 \\ i<j<k}}^{m} \sum_{j=1}^{m} \sum_{k=1}^{m} n\left(A_i \cap A_j \cap A_k\right) - \dots$$

$$\dots . + (-1)^{m-1} n\left(\bigcap_{i=1}^{m} A_i\right)$$

where $n(A)$ represents the number of elements in A.

Note:

1. $n(A_1 \cup A_2) = n(A_1) + n(A_2) - n(A_1 \cap A_2)$
2. $n(A_1 \cup A_2 \cup A_3) = n(A_1) + n(A_2) + n(A_3) - n(A_1 \cap A_2) - n(A_1 \cap A_3)$

$$-n(A_2 \cap A_3) + n(A_1 \cap A_2 \cap A_3)$$

Example 1: Find the number of ways of selecting

(i) A diamond or heart

(ii) An ace or a spade

from a pack of 52 cards

Solution: Let A_1 be the set of diamonds, A_2 be the set of hearts, A_3 be the set of aces and A_4 set of spades.

(i) Here $(A_1) = 13$, $n(A_2) = 13$ and A_1, A_2 are disjoint.
Hence $n(A_1 \cap A_2) = 0$ and
$n(A_1 \cup A_2) = n(A_1) + n(A_2) - n(A_1 \cap A_2) = 13 + 13 - 0 = 26$

(ii) Here $n(A_3) = 4$ and $n(A_4) = 13$. Note that A_3 and A_4 are not disjoint and $n(A_3 \cap A_4) = 1$. Hence $n(A_3 \cup A_4) = 4 + 13 - 1 = 16$

Note: If $A_1, A_2, \dots, A_n$ are pair-wise disjoint sets, then there will be no common elements to these sets and hence

$$n\left(\bigcup_{i=1}^{m} A_i\right) = \sum_{i=1}^{m} n(A_i)$$

Fundamental Principle of Multiplication (Product rule)

Let $A_1, A_2, \dots, A_m$ be m sets and the elements in each set are distinct. Then the number of ways of selecting first object from A_1, second object from A_2, …, m^{th} object from A_m in succession is given by

$$n(A_1 \times A_2 \times \dots \times A_m) = n(A_1).n(A_2)\dots n(A_m)$$

Example 2: A man has 8 different shirts and 6 different pants. In how many different ways, he can be dressed?

Solution: Choosing a dress means selection of one shirt and one pant. The total number of ways of choosing a dress is $8 \times 6 = 48$

Example 3: Two dice are thrown.

(i) How many different outcomes are there?
(ii) How many different outcomes with distinct values (no doubles)?

Solution: On each die, we may get the number 1 or 2 or 3 or 4 or 5 or 6.
One outcome means one number on first die and another number on second die.

(i) Number of different out comes $= 6 \times 6 = 36$
(ii) Number of different out comes $= 6 \times 5 = 30$

Permutations

A permutation is an arrangement or an ordered selection of objects. There is importance to the order of objects in a permutation.

1) The number of permutations of n different objects taken $r(\leq n)$ at a time is

   ${}^{n}P_{r} = n(n-1)\ldots\big(n-(r-1)\big)$ when repetition of objects is not allowed.

   The number of permutations of n different objects taken $r(\leq n)$ at a time is

   n^r when repetition of objects is allowed any number of times.

2) The number of permutations of n different objects taken all at a time when repetition of objects is not allowed is $n!$
3) If there are n objects, n_1 of type 1, n_2 of type 2,..., n_k of type k , where $n_1+n_2+n_3\ldots+n_k=n$, then the number of permutations of these n objects taken all at a time is

$$\frac{n!}{n_1!\,n_2!\,n_3!\ldots n_k!}$$

4) The number of permutations of n different objects taken r at a time without repetitions in which
   - (i) k particular objects will always occur is ${}^{n-k}P_{r-k}\ {}^{r}P_{k}$
   - (ii) s particular objects will never occur is ${}^{(n-s)}P_{r}$
   - (iii) k particular objects will always occur and s particular objects will never occur is ${}^{(n-k-s)}P_{r-k}\ {}^{r}P_{k}$

Combinations

A combination is an unordered selection or subset of the objects. There is no importance to the order of the objects in a combination.

1) The number of combinations of n different objects taken $r(\leq n)$ at a time is

   denoted by ${}^{n}C_{r}$ and ${}^{n}C_{r}=\dfrac{{}^{n}P_{r}}{r!}$ when repetition of objects is not allowed.

The number of combinations of n different objects taken $r(\leq n)$ at a time is $(n+r-1)_{C_r}$ when repetition of objects is allowed.

2) The number of combinations of n different objects taken r at a time without repetitions in which
   (i) k particular objects will always occur is $(n-k)_{C_{r-k}}$
   (ii) s particular objects will never occur is $(n-s)_{C_r}$
   (iii) k particuarl objects will always occur and s particular objects will never occur is $(n-k-s)_{C_{r-k}}$
3) The number of combinations of n different objects taken any number (one or more) at a time when repetitions are not allowed is

$$n_{C_1} + n_{C_2} + \cdots + n_{C_n} = 2^n - 1$$

4) The total number of combinations of $(n_1 + n_2 + \cdots + n_k)$ objects taken any number at a time when n_1 objects are of type 1, n_2 are of type 2, … , n_k are of type $k = (n_1 + 1)(n_2 + 1) \ldots (n_k + 1) - 1$

5) The total number of combinations of $(n_1 + n_2 + \cdots + n_k + m)$ objects **taken any number at a time** when n_1 objects are of type 1, n_2 are of type 2, … , n_k are of type k =$(n_1 + 1)(n_2 + 1) \ldots (n_k + 1). 2^m - 1$

Circular Permutations

An arrangement of objects arranged in a circle is known as a circular permutation.

1) The number of circular permutations of n different objects taken all at a time is $(n-1)!$
2) The number of circular permutations of n different objects taken all at a time when clockwise and anticlockwise arrangements are considered the same (as in Necklace, Garland) is $\frac{(n-1)!}{2}$

3) The number of circular permutations of n different objects taken r at a time is $n_{C_r}(r-1)! = \frac{n_{P_r}}{r}$

4) The number of circular permutations of n different objects taken r at a time when no distinction is made between clockwise and anticlockwise direction $= \frac{1}{2} \cdot \frac{n_{P_r}}{r}$.

Distribution or Occupancy Problems

The number of ways, r objects can be distributed among n different boxes, depends upon the fact: how many objects are permitted to be in one box and whether the objects are different or not. Problems involving the distribution of objects among boxes are called distribution or occupancy problems.

The distribution of different objects corresponds to permutations and distribution of identical objects corresponds to combinations.

Distribution of Different Objects:

1. The number of ways of distributing r different objects into n different boxes if

(i) no restriction is placed on the number of objects permitted in a box is n^r.

(ii) a particular box contains exactly k objects is $r_{C_k} \cdot (n-1)^{r-k}$.

(iii) at most one object is permitted into a box is $n_{P_r} (n \geq r)$.

2. The number of ways of distributing r_i objects to the i^{th} box for $i = 1, 2, \ldots, n$ such that $r_1 + r_2 + \cdots + r_n = r$ is given by

$$\frac{r\,!}{r_1\,!\, r_2\,!\, r_3\,! \ldots r_n\,!}$$

Distribution of Identical objects

1) The number of ways of distributing r identical objects into n different boxes if

(i) no restriction is placed on the number of objects permitted per box is $(n+r-1)_{C_r}$ (*Bose – Einstein formula*)

(ii) A particular box contains exactly k objects is

$$\left((n-1)+(r-k)-1\right)_{C_{r-k}} = (n-r-k-2)_{C_r}$$

(iii) atmost one object is permitted per box is $n_{C_r}\ (r \le n)$

(*Fermi – Dirac formula*)

Example 4

S.No	**Objects**	**Arrangement**	**Problem**	**Answer**
1	5 boys and 4 girls	Row	No two girls together	$5! \times 6_{P_4}$
2	5 boys and 4 girls	Circle	No two girls together	$4! \times 5_{P_4}$
3	5 boys and 5 girls	row	No two girls together	$5! \times 6_{P_5}$
4	5 boys and 5 girls	row	Boys and girls alternate	$5! \times 5! \times 2!$
5	5 boys and 5 girls	circle	No two girls together	$4! \times 5_{P_5} = 4! \times 5!$
6	5 boys and 5 girls	circle	Boys and girls alternate	$4! \times 5!$
7	$5 +$ signs and $4 -$ signs	row	No two $-s$ together	$1 \times 6_{C_4}$
8	$5 +$ signs and $4 -$ signs	circle	No two $-s$ together	$1 \times 5_{C_4}$
9	$5 +$ signs and $5 -$ signs	row	No two $-s$ together	$1 \times 6_{C_5}$
10	$5 +$ signs and $5 -$ signs	row	$+$ and $-$ alternate	$1 + 1 = 2$
11	$5 +$ signs and $5 -$ signs	circle	No two $-s$ together	1
12	$5 +$ signs and $5 -$ signs	circle	$+$ and $-$ alternate	1

Example 5:

(i) Find the number of 4 -letter words that can be formed using the letters of the word **EQUATION.**

(ii) How many of these words begin with E?

(iii) How many end with N?

(iv) How many begin with E and end with N?

Solution: The word EQUATION has 8 distinct letters.

(i) Number of 4 letter words is $8_{\boldsymbol{P}_4}$

(ii) The first letter E is fixed $(E - --)$. The remaining three letters are to be filled with 7 letters. Thus, the number of $4 -$ letter words begin with E is $7_{\boldsymbol{P}_3}$

(iii) $- - -N$. No of words ending with N is $7_{\boldsymbol{P}_3}$

(iv) $E - -N$. No of words begin with E and end with N is $6_{\boldsymbol{P}_2}$

Example 6: Find the number of 4 letter words that can be formed using the letters of the word **MIXTURE** which

(i) contain the letter X

(ii) do not contain the letter X

Solution: Take 4 blanks $- - - -$. We have to fill up 4 blanks using the 7 letters of the word.

(i) First we put X in one of the 4 blanks. This can be done in 4 ways. Now we can fill the remaining 3 palces with the remaining 6 letters in 6_{P_3} ways. Thus, the number of 4 letter words containing the letter X are $4 \times 6_{P_3} = 4 \times 120 = 480$

(ii) Leaving the letter X, we fill the 4 blanks with the remaining 6 letters in 6_{P_4} ways. Thus, the number of 4 letter words that do not contain the letter X is

$$6_{P_4} = 360$$

Example 7: Find all 4 –digit numbers that can be formed using the digits $1, 2, 3, 4, 5, 6$ when repetition is allowed.

Solution: The number of 4 –digit number with repetitions is 6^4

Example 8: Find the number of ways of arranging the letters of the word **SPECIFIC**. In how many of them

(i) the two Cs come together?
(ii) the two Is do not come together?

Solution: The word **SPECIFIC** has 8 letters in which there are 2 I's and 2 C's. Hence, they can be arranged in

$$\frac{8!}{2!2!} \text{ ways}$$

(i) Treat two C's as one unit. Then we have $6 + 1 = 7$ letters in which two letters (I's) are alike.

Thus, the no. of arrangement $= \frac{7!}{2!}$

(ii) Keeping the two I's aside, arrange the remaining 6 letters can be arragned in $\frac{6!}{2!}$ ways. Among these 6 letters we find 7 gaps as shown below.

$$-S - P - E - C - F - C -$$

The two I's can be arranged in these 7 gaps in $\frac{{}^7\boldsymbol{P}_2}{2!}$

Hence, the number of required arrangements is $\frac{6!}{2!} \times \frac{{}^7\boldsymbol{P}_2}{2!}$

Example 9: Find the number of ways of selecting 4 boys and 3 girls from a group of 8 boys and 5 girls is

Solution: $8_{C_4} \times 5_{C_3}$

Example 10: Find the number of ways of forming a committee of 4 members out of 6 boys and 4 girls such that there is at least one girl in the committee.

Solution: $10_{C_4} - 6_{C_4}$

Derangements and Matches

If n objects numbered $1, 2, 3, \dots, n$ are distributed at random in n places also numbered $1, 2, \dots, n$ a match is said to occur. If an object occupies the place corresponding to its number, the number of permutations in which no match occurs is

$$\boldsymbol{D_n} = \boldsymbol{n!}\left\{1 - \frac{1}{1\boldsymbol{!}} + \frac{1}{2\boldsymbol{!}} - \frac{1}{3\boldsymbol{!}} + \boldsymbol{\dots} + (-1)^{\boldsymbol{n}}\frac{1}{\boldsymbol{n!}}\right\}$$

This is also known as derangement.

The number of permutations of n objects in which exactly r matches occur is

$$n_{C_r}.D_{n-r} = \frac{n!}{r!}\left\{1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \cdots + \frac{(-1)^{n-r}}{(n-r)!}\right\}$$

1.2

Basic Concepts in Probability

Introduction to uncertainty

Every day we have been coming across statements like the ones mentioned below:

1. Probably it will rain tonight.
2. It is quiet likely that there will be a good yield of paddy this year.
3. Probably I will get a first class in the examination.
4. India might win the cricket series against Australia
   and so on.

In all the above statements some element of uncertainty or chance is involved. A numerical measure of uncertainty is provided by a very important branch of statistics known as **Theory of Probability**. In the words of *Prof. Ya-Lin-Chou*: *Statistics is the science of decision making with calculated risks in the face of uncertainty*.

History of Probability

The history of probability suggests that its theory developed with the study of *games of chance*, such as *rolling of dice*, *drawing a card from a pack of cards*, etc. Two French gamblers had once decided that any one person who will first get a 'particular point' will win the game. If the game is stopped before reaching that point, the question is how to share the stake. This and similar other problems were then posed by the great French mathematician *Blaise Pascal*, who after consulting another great French mathematician *Pierre de Fermat*, gave the solution of the problems and then laid down a strong foundation of probability. Later on, another French mathematician, *Laplace*, improved the definition of probability.

Coins, Dice and Playing Cards: The basic concepts in probability are better explained using *coins*, *dice* and *playing cards*. The knowledge of these is very much useful in solving problems in probability.

Coin: A coin is round in shape and it has two sides. One side is known as ***head (H)*** and the other is known as ***tail (T).*** When a coin is tossed, the side on the top is known as the result of the toss.

Die: A die is cube in shape in which length, breadth and height are equal. It has six faces which have same area and numbered from 1 to 6. The plural of die is dice. When a die is thrown, the number on the top face is the result of the throw.

Pack of Cards: A pack of cards 52 cards. It is divided into four suits called *spades*, *clubs*, *hearts* and *diamonds*. Spades and clubs are black; hearts and diamonds are red in colour. Each suit consists of 13 cards, of which *nine* cards are numbered from 2 to 10, an ace, jack, queen and king. We shuffle the cards and then take a card from the top which is the result of selecting a card.

Basic Concepts in Probability

The following basic concepts are very important in understanding the definitions of the probability:

Experiment: The process of making an observation or measurement and observation about a phenomenon is known as an ***experiment***.

Example1: Sitting in the balcony of the house and watching the movement of clouds in the sky is an experiment.

Example2: For given values of pressure (P), measuring the corresponding values of volume (V) of a gas and observing that $P \cdot V = k$(constant) is an experiment. The experiments are of two types:

Deterministic experiment: If an experiment produces the same result when it is conducted several times under identical conditions, then the experiment is known as ***determinant experiment***.
All the experiments in physical and engineering sciences are deterministic.

Random Experiment: If an experiment produces different results even though it is conducted several times under identical conditions, then the experiment is known as ***random experiment***. All the experiments in social sciences are random.

Trial: Conducting a random experiment once is known as a ***trial***.

Outcome: A result of a random experiment in a trial is known as an ***outcome***.
Outcomes are denoted by lowercase letters $a, b, c, d, e, \dots$.

Equally Likely Outcomes: Outcomes of a random experiment are said to be ***equally likely*** if all have the same chance of occurrence. Getting a H and T in a balanced coin are equally likely. The outcomes $1{,}2{,}3{,}4{,}5$ and 6 are equally likely if the die is a cube.

Sample space: The set of all possible outcomes of a random experiment is known as a ***sample space*** and denoted by **S**.

Event: A subset of the sample space is known as an ***event***.
The events are denoted by uppercase letters $\mathrm{A}, \mathrm{B}, \mathrm{C}$ etc.

Happening of an event: We say that an event happens (or occurs) if any one outcome in it happens (or occurs).

Elementary Event: A singleton set consisting an outcome of a random experiment is known as an ***elementary event***.

Favorable outcomes: The outcomes in an event are known as ***favorable outcomes*** or ***cases*** of that event.

Impossible Event: An event with no outcome in it is known as ***impossible event*** and is denoted by $\boldsymbol{\phi}$.

Certain or Sure Event: An event consisting of all possible outcomes of a random experiment is known as ***certain*** or ***sure event*** and it is same as the sample space.

Exhaustive Events: The events in a sample space are said to be ***exhaustive*** if their union is equal to the sample space. The events $A_1, A_2, \dots, A_n$ in S are said to be exhaustive if

$$\bigcup_{i=1}^{n} A_i = S$$

Mutually Exclusive Events: Two or more events in the sample space are said to be ***mutually exclusive*** if the happening of one of them precludes the happening of the others. Mathematically two events A and B in S are said to be mutually exclusive if $A \cap B = \phi$.

Example3: Consider a random experiment of tossing a coin. The possible outcomes are H and T. Thus, the sample space is given by $S = \{H, T\}$ and $n(S) = 2$ where $n(S)$ is the total number of outcomes in S.

Example 4: Consider a random experiment of tossing two coins (or two tosses of a coin). The sample space is given by $S = \{H, T\} \times \{H, T\} = \{HH, HT, TH, TT\}$ and $n(S) = 2^2 = 4$.

Example 5: Consider a random experiment of tossing three coins (or three tosses of a coin). The sample space is given by

$$S = \{H, T\} \times \{H, T\} \times \{H, T\} = \{H, T\} \times \{HH, HT, TH, TT\}$$

$$= \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$

and $n(S) = 2^3 = 8$.

Let us define some events in the sample space as below:

E_1: Three heads

E_2: Three tails

E_3: Exactly one head

E_4: Exactly two heads

E_5: At least one head

E_6: At least two heads

Then these events are represented by the following subsets of S:

$E_1 = \{HHH\}$;

$E_2 = \{TTT\}$

$E_3 = \{HTT, THT, TTH\}$;

$E_4 = \{HHT, HTH, THH\}$;

$E_5 = \{HHH, HHT, HTH, HTT, THH, THT, TTH\}$ and
$E_6 = \{HHH, HHT, HTH, THH\}$.

Note that $E_1 \cup E_2 \cup E_3 \cup E_4 = S$ and hence E_1, E_2 ,E_3 andE_4 are exhaustive events in S. Further, $E_i \cap E_j = \phi$, where $i \neq j$. Hence, E_1, E_2 ,E_3 and E_4 are mutually exclusive events in S.

Note: In general, if a random experiment consists of tossing N coins (or N tosses of a coin), then $n(S) = 2^N$.

Example 6: Let us consider a random experiment of throwing a die. Since we can obtain any one of the six faces 1,2,3,4,5 and 6, the sample space is given by $S = \{1,2,3,4,5,6\}$ and $n(S) = 6$.

Now define $E_1 = \{1,3,5\}, E_2 = \{2,4,6\}$ and $E_3 = \{3,6\}$. We say that E_1 happens or occurs if we get the outcome $1,3$ or 5. In otherwords, we say that E_1 happens if we get an odd number. Similarly, we say that E_2 happens if we get an even number and E_3 happens if we get a multiple of 3.

Since E_1, E_2 and E_3 are subsets of S; E_1, E_2 and E_3 are events in S. Since $E_1 \cup E_2 = S$, E_1 and E_2 are exhaustive events in S. Since $E_1 \cup E_3 = \{1,3,5,6\} \neq S$, E_1 and E_3 are not exhaustive events in S. Since $E_1 \cap E_2 = \phi$, E_1 and E_2 are mutually exclusive events in S. Since $E_1 \cap E_3 = \{3\}$, E_1 and E_3 are not mutually exclusive events in S. Similarly E_2 and E_3 are not mutually exclusive events in S.

Example 7: In a random experiment of throwing two dice (or two throws of a die), the sample space is given by

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

$$\{(1,1),(1,2),\dots,(1,6),$$

$$(2,1),(2,2),\dots,(2,6),$$

$$(3,1),(3,2),\dots,(3,6)$$

$$(4,1),(4,2),\dots,(4,6)$$

$$(5,1),(5,2),\dots,(5,6)$$

$$(6,1),(6,2),\dots,(6,6)\}$$

where in the outcome (a,b), a represents the number obtained on the first die and b represents the number on the second die. Obviously $(a,b) \neq (b,a)$ unless $a = b$. The number of outcomes in S is given by $S = 6^2 = 36$.

Let us define the following events in S.

E_1: Sum of points on two dice is 5

E_2: Sum of points on two dice is 6

E_3: Sum of points on two dice is even

E_4: Sum of points on two dice is odd

E_5: Sum of points on two dice is greater than 12

E_6: Sum of points on two dice is divisible by 3

E_7: Sum is greater than or equal to 2 and is less than or equal to 12

Then the events E_1 to E_7 as subsets of S are given below.

$E_1 = \{(1,4), (2,3), (3,2), (4,1)\}$ and $n(E_1) = 4$

$E_2 = \{(1,5), (2,4), (3,3), (4,2), (5,1)\}$ and $n(E_2) = 5$

The sum of the points on the two dice is even if the points obtained on each die is (i) even or (ii) odd. Thus

$E_3 = (\{2,4,6\} \times \{2,4,6\}) \cup (\{1,3,5\} \times \{1,3,5\})$

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

and $n(E_3) = (3 \times 3) + (3 \times 3) = 9 + 9 = 18.$

Similarly,

$E_4 = (\{2,4,6\} \times \{1,3,5\}) \cup (\{1,3,5\} \times \{2,4,6\})$

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

and $n(E_4) = (3 \times 3) + (3 \times 3) = 18.$

Further, $E_5 = \phi, i.e.,$ E_5 is an impossible event and $E_7 = S, i.e.,$ E_7 is a certain event. Hence $n(E_5) = 0$ and $n(E_7) = 36$.

The sum of the points on the two dice is divisible by 3 if their sum is 3,6,9 or 12. Thus

$E_6 = \{(1,2), (2,1), (1,5), (2,4), (3,3), (4,2), (5,1), (3,6), (4,5), (5,4), (6,3), (6,6)\}$ and $n(E_6) = 12$.

Note: In general, if the random experiment consists of throwing of N dice (or N throws of a die), the number of outcomes in S is given by $n(S) = 6^N$.

Example 8: Let us consider the random experiment of tossing a coin and a die together. Then the sample space is given by

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

$= \{(H, 1), (H, 2), (H, 3), (H, 4), (H, 5), (H, 6), (T, 1), (T, 2), (T, 3), (T, 4), (T, 5), (T, 6)\}$

and $n(S) = 2 \times 6 = 12$.

Note: In the above examples 3 to 8, if the coins and dice are unbiased, the outcomes in the sample spaces are equally likely. Normally, the coins are balanced and hence are unbiased. If a die is a cube, then all the surfaces have the same area and also it is unbiased.

Example 9: Let us consider the random experiment of selecting two balls simultaneously from and urn containing 4 balls of different colours red(R), blue(B), yellow(Y) and white(W). Then the sample space is given by

$$S = \{RB, RY, RW, BY, BW, YW\} \text{ and } n(S) = 4_{C_2} = 6$$

Example 10: If the random experiment consists of selecting two balls one after the other with replacement in Example 9, the sample space is given by

$S = \{R, B, Y, W\} \times \{R, B, Y, W\} =$
$\{RR, RB, RY, RW, BR, BB, BY, BW, YR, YB, YY, YW, WR, WB, WY, WW\}$ and $n(S) = 4 \times 4 = 16$.

Example 11: If the random experiment consists of selecting two balls one after the other without replacement in Example7, the sample space is given by

$S = \{RB, RY, RW, BR, BY, BW, YR, YB, YW, WR, WB, WY\}$ and
$n(S) = 4 \times 3 = 12$.

Example12: Consider a random experiment of tossing a coin until head appears. Its sample space is given by

$$S = \{H, TH, TTH, TTTH, \dots\}$$

where TTH represents tail in first, tail in second and head in third tosses and so on. Obviously, $n(S)$ is infinite.

Example13: Consider a random experiment of tossing a coin repeatedly until head or tail appears twice in succession. Thus the sample space is given by

$$S = \{HH, TT, THH, HTT, HTHH, THTT, \dots\}$$

and $n(S)$ is infinite.

1.3.

Definitions of Probability

The probability of a given event is an expression of likelihood or chance of occurrence of an event. How the number is assigned would depend on the interpretation of the term 'probability'. There is no general agreement about its interpretation. However, broadly speaking, there are four different schools of thought on the concept of probability.

Mathematical (or classical or A priori) definition of probability

Let S be a sample space associated with a random experiment. Let A be an event in S. We make the following assumptions on S:

(i) It is discrete and finite

(ii) The outcomes in it are equally likely

Then the probability of happening (or occurrence) of the event A is defined by

$$P(A) = \frac{Number\ of\ outcomes\ in\ A}{Number\ of\ outcomes\ in\ S} = \frac{n(A)}{n(S)}$$

Note:

i) The probability of non-happening (or non-occurrence) of A is given by

$$P(\bar{A}) = \frac{Number\ of\ outcomes\ in\ \bar{A}}{Number\ of\ outcomes\ in\ S} = \frac{n(\bar{A})}{n(S)} = \frac{n(S)-n(A)}{n(S)} = 1 - \frac{n(A)}{n(S)} = 1 - P(A)$$

That is $P(\bar{A}) = 1 - P(A)$

ii) If $A = \phi$, then $P(\phi) = \frac{n(\phi)}{n(S)} = \frac{0}{n(S)} = 0$. That is, probability of an impossible event is zero.

iii) If $A = S$, then $P(S) = \frac{n(S)}{n(S)} = 1$. That is, probability of a certain event is one.

iv) For any event A in S, $0 \leq P(A) \leq 1$.

v) The odds in favour of A are given by $n(A) : n(\bar{A}) = P(A) : P(\bar{A})$.

vi) The odds against of A are given by $n(\bar{A}) : n(A) = P(\bar{A}) : P(A)$.

vii) If the odds in favour of A are $a : b$, then $P(A) = \frac{a}{a+b}$.
viii) If the odds against of A are $c : d$, then $P(A) = \frac{d}{c+d}$.
ix) $n(A)$ and $n(S)$ are counted by using methods of counting discussed in **Module 1.1**.

Limitations: The mathematical definition of probability breaks down in the following cases:

(i) The outcomes in the sample space are not equally likely.
(ii) The number of outcomes in the sample space is infinite.

Statistical (or Empirical or Relative Frequency or Von Mises) Definition of Probability

If a random experiment is performed repeatedly under identical conditions, then the limiting value of the ratio of the number of times the event occurs to the number of trials, as the number of trials becomes indefinitely large, is called the probability of happening of the event, it being assumed that the limit is finite and unique.

Symbolically, if in N trials an event A happens a_N times, then the probability of the happening of A is given by

$$P(A) = \lim_{N \to \infty} \frac{a_N}{N} \qquad \dots (1.3.1)$$

Note:

i) Since the probability is obtained objectively by repetitive empirical observations, it is known as Empirical Probability.
ii) The empirical probability approaches the classical probability as the number of trials becomes indefinitely large.

Limitations of Empirical Probability

(i) If an experiment is repeated a large number of times, the experimental conditions may not remain identical.
(ii) The limit in (1.3.1) may not attain a unique value, however large N may be.

Subjective definition of probability: In this method, probabilities are assigned to events according to the knowledge, experience and belief about the happening of the events. The main limitation of this definition is, it varies from person to person.

Axiomatic Definition of Probability: Let S be a sample space and let $\mathbb{B}$ be a σ-field associated with S. A probability function (or measure) P is a real valued set function having domain B and which satisfies the following three axioms:

1. $P(A) \geq 0$, for every $A \in \mathbb{B}$ (Non-negativity)
2. $P(S) = 1$, $i.e$, P is normed (Normality)
3. If $A_1, A_2, \dots, A_n, \dots$ are mutually exclusive events in S, then
   $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$ (σ - additive or countably additive)

Thus, the probability function is a normed measure on (the measurable space) $(\mathrm{S}, \mathbb{B}, \mathrm{P})$ is called a **Probability space**. This definition is useful in proving theorems on probability.

Note: The elements of $\mathbb{B}$ are events in S.

Solved Examples using Mathematical Definition of Probability

In this section, we use mathematical definition of probability for computing probabilities. Also we use methods of counting for counting the number of outcomes in an event and sample space.

Example 1: A uniform die is thrown at random. Find the probability that the number on it is (i) even (ii) odd (iii) even or multiple of 3 (iv) even and multiple of 3 (v) greater than 4

Solution:

(i) The number of favourable cases to the event of getting an even number is 3, viz., 2,4,6.
$\therefore$ Required probability $= \frac{3}{6} = \frac{1}{2}$

(ii) The number of favourable cases to the event of getting an odd number is 3, viz., 1, 3, 5.

$\therefore$ Required probability $= \frac{3}{6} = \frac{1}{2}$

(iii) The number of favourable cases to the event of getting even or multiple of 3 is 4, viz., 2, 3, 4, 6.

$\therefore$ Required probability $= \frac{4}{6} = \frac{2}{3}$

(iv) The number of favourable cases to the event of getting even and multiple of 3 is 1, viz., 6.

$\therefore$ Required probability $= \frac{1}{6}$

(v) The number of favourable cases to the event of getting greater than 4 is 2, viz., 5 and 6.

$\therefore$ Required probability $= \frac{2}{6} = \frac{1}{3}$

Example 2: Four cards are drawn at random from a pack of 52 cards. Find the probability that

(i) They are a king, a queen, a jack and an ace.
(ii) Two are kings and two are aces.
(iii) All are diamonds.
(iv) Two are red and two are black.
(v) There is one card of each suit.
(vi) There are two cards of clubs and two cards of diamonds.

Solution: Four cards can be drawn from a well shuffled pack of 52 cards in ${}^{52}C_4$ ways, which gives the exhaustive number of cases.

(i) 1 king can be drawn out of the 4 kings is ${}^{4}C_1 = 4$ ways. Similarly, 1 queen, 1 jack and an ace can each be drawn in ${}^{4}C_1 = 4$ ways. Since any one of the ways of drawing a king can be associated with any one of the ways of drawing a queen, a jack and an ace, the favourable number of cases are ${}^{4}C_1 \times {}^{4}C_1 \times {}^{4}C_1 \times {}^{4}C_1$.

Hence, required probability $= \frac{{}^{4}C_1 \times {}^{4}C_1 \times {}^{4}C_1 \times {}^{4}C_1}{{}^{52}C_4} = \frac{256}{{}^{52}C_4}$

(ii) Required probability $=\frac{{}^{4}C_2\times{}^{4}C_2}{{}^{52}C_4}$

(iii) Since 4 cards can be drawn out of 13 cards (since there are 13 cards of diamond in a pack of cards) in ${}^{13}C_4$ ways,

Required probability $=\frac{{}^{13}C_4}{{}^{52}C_4}$

(iv) Since there are 26 red cards (of diamonds and hearts) and 26 black cards (of spades and clubs) in a pack of cards,

Required probability $=\frac{{}^{26}C_2\times{}^{26}C_2}{{}^{52}C_4}$

(v) Since, in a pack of cards there are 13 cards of each suit,

Required probability $=\frac{{}^{13}C_1\times{}^{13}C_1\times{}^{13}C_1\times{}^{13}C_1}{{}^{52}C_4}$

(vi) Required probability $=\frac{{}^{13}C_2\times{}^{13}C_2}{{}^{52}C_4}$

Example 3: What is the chance that a non-leap year should have fifty-three Sundays?

Solution: A non-leap year consists of 365 days, $i.e.$, 52 full weeks and one over-day. A non-leap year will consist of 53 Sundays if this over-day is Sunday. This over-day can be anyone of the possible outcomes:

(i) Sunday (ii) Monday (iii) Tuesday (iv) Wednesday (v) Thursday (vi) Friday (vii) Saturday, $i.e.$, 7 outcomes in all. Of these, the number of ways favourable to the required event viz., the over-day being Sunday is 1.

$\therefore$ Required probability $=\frac{1}{7}$

Example 4: Find the probability that in 5 tossings, a perfect coin turns up head at least 3 times in succession.

Solution: In 5 tossings of a coin, the sample space is:

$$S = \{H,T\} \times \{H,T\} \times \{H,T\} \times \{H,T\} \times \{H,T\}, (H : \text{head} ; T : \text{tail})$$

$\therefore$ Exhaustive number of cases $= 2^5 = 32$.

The favourable cases for getting at least three heads in succession are :

Starting with 1st toss: $HHHTH$, $HHHTT$, $HHHHT$, $HHHHH$

Starting with 2nd toss: $THHHT$, $THHHH$

Starting with 3rd toss: $TTHHH$, $HTHHH$

Hence, the total number of favourable cases for getting at least 3 heads in succession are 8.

$\therefore$ Required probability $= \frac{Number\ of\ favourable\ cases}{Exhaustive\ number\ of\ cases} = \frac{8}{32} = \frac{1}{4} = 0.25$

Example 5: A bag contains 20 tickets marked with numbers 1 to 20. One ticket is drawn at random. Find the probability that it will be a multiple of (i)2 or 5, (ii)3 or 5

Solution: One ticket can be drawn out of 20 tickets in ${}^{20}C_1 = 20$ ways, which determine the exhaustive number of cases.

(i) The number of cases favourable to getting the ticket number which is:

(a) a multiple of 2 are 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, $i.e.$, 10 cases.

(b) a multiple of 5 are 5, 10, 15, 20 i.e., 4 cases

Of these, two cases viz., 10 and 20 are duplicated.

Hence the number of distinct cases favourable to getting a number which is a multiple of 2 or 5 are: $10 + 4 - 2 = 12$.

$\therefore$ Required probability $= \frac{12}{20} = \frac{3}{5} = 0.6$

(ii) The cases favourable to getting a multiple of 3 are 3, 6, 9, 12, 15, 18 i.e., 6 cases in all and getting a multiple of 5 are 5, 10, 15, 20 i.e., 4 cases in all. Of these, one case viz., 15 is duplicated.

Hence, the number of distinct cases favourable to getting a multiple of 3 or 5 is $6 + 4 - 1 = 9$.

$\therefore$ Required probability $= \frac{9}{20} = 0.45$

Example 6: An urn contains 8 white and 3 red balls. If two balls are drawn at random, find the probability that

(i) both are white, (ii) both are red, (iii) one is of each color.

Solution: Total number of balls in the urn is $8 + 3 = 11$. Since 2 balls can be drawn out of 11 balls in ${}^{11}C_2$ ways,

Exhaustive number of cases $= {}^{11}C_2 = \frac{11\times10}{2} = 55$

(i) If both the drawn balls are white, they must be selected out of the 8 white balls and this can be done in ${}^{8}C_2 = \frac{8\times7}{2} = 28$ ways.

$\therefore$ Probability that both the balls are white $= \frac{28}{55}$

(ii) If both the drawn balls are red, they must be drawn out of the 3 red balls and this can be done in ${}^{3}C_2 = 3$ ways. Hence, the probability that both the drawn balls are red $= \frac{3}{55}$.

(iii) The number of favourable cases for drawing one white ball and one red ball is ${}^{8}C_1 \times {}^{3}C_1 = 8\times3 = 24$

$\therefore$ Probability that one ball is white and other is red $= \frac{24}{55}$

Example 7: The letters of the word 'article' are arranged at random. Find the probability that the vowels may occupy the even places.

Solution: The word 'article' contains 7 distinct letters which can be arranged among themselves in $7!$ ways. Hence exhaustive number of cases is $7!$.

In the word 'article' there are 3 vowels, viz., a, i and e and these are to be placed in, three even places, viz., 2nd, 4th and 6th place. This can be done in $3!$, ways. For each arrangement, the remaining 4 consonants can be arranged in $4!$ ways. Hence, associating these two operations, the number of favourable cases for the vowels to occupy even places is $3! \times 4!$.

$\therefore$ Required probability $= \frac{3!4!}{7!} = \frac{3!}{7\times6\times5} = \frac{1}{35}$

Example 8: Twenty books are placed at random in a shelf. Find the probability that a particular pair of books shall be:

(i) Always together **(ii) Never together**

Solution: Since 20 books can be arranged among themselves in $20!$ ways, the exhaustive number of cases is $20!$.

(i) Let us now regard that the two particular books are tagged together so that we shall regard them as a single book. Thus, now we have $(20-1) = 19$ books which can be arranged among themselves in $19!$ ways. But the two books which are fastened together can be arranged among themselves in $2!$ ways.

Hence, associating these two operations, the number of favourable cases for getting a particular pair of books always together is $19! \times 2!$.

$\therefore$ Required probability is $\frac{19!\times2!}{20!} = \frac{2}{20} = \frac{1}{10}$.

(ii) Total number of arrangement of 20 books among themselves is $20!$ and the total number of arrangements that a particular pair of books will always be together is $19!\,2!$, [See part (i)]. Hence, the number of arrangements in which a particular pair of books is never together is

$$20! - 2 \times 19! = (20-2) \times 19! = 18 \times 19!$$

$\therefore$ Required probability $= \frac{18 \times 19!}{20!} = \frac{18}{20} = \frac{9}{10}$

Aliter: P [A particular pair of books shall never be together]

$= 1 - P$[A particular pair of books is always together]$= 1 - \frac{1}{10} = \frac{9}{10}$.

Example 9: n persons are seated on n chairs at a round table. Find the probability that two specified persons are sitting next to each other.

Solution: The n persons can be seated in n chairs at a round table in $(n-1)!$ ways, which gives the exhaustive number of cases.

If two specified persons, say, A and B sit together, then regarding A and B fixed together, we get $(n-1)$ persons in all, who can be seated at a round table in $(n-2)!$ ways. Further, since A and B can interchange their positions in 2! ways, total number of favourable cases of getting A and B together is $(n-2)! \times 2!$. Hence, the required probability is: $\frac{(n-2)! \times 2!}{(n-1)!} = \frac{2}{n-1}$

Aliter: Let us suppose that of the n persons, two persons, say, A and B are to be seated together at a round table. After one of these two persons, say A occupies the chair, the other person B can occupy any one of the remaining $(n-1)$ chairs. Out of these $(n-1)$ seats, the number of seats favourable to making B sit next to A is 2 (since B can sit on either side of A). Hence the required probability is $\frac{2}{n-1}$.

Example 10: In a village of 21 inhabitants, a person tells a rumour to a second person, who in turn repeats it to a third person, etc. at each step the recipient of the rumour is chosen at random from the 20 people available. Find the probability that the rumour will be told 10 times without:

(i) returning to the originator ; **(ii) being repeated to any person**

Solution: Since any person can tell the rumour to any one of the remaining $21 - 1 = 20$ people in 20 ways, the exhaustive number of cases that the rumour will be told 10 times is 20^{10}.

(i) Let us define the event :

E_1 : The rumour will be told 10 times without returning to the originator.

The originator can tell the rumour to any one of the remaining 20 persons in 20 ways, and each of the $10 - 1 = 9$ recipients of the rumour can tell it to any of the remaining $20 - 1 = 19$ persons (without returning it to the originator) in 19 ways. Hence the favourable number of cases for E_1 are 20×19^9. The required probability is given by :

$$P(E_1) = \frac{20 \times 19^9}{20^{10}} = \left(\frac{19}{20}\right)^9$$

(ii) Let us define the event :

E_2 : The rumour is told 10 times without being repeated to any person.

In this case the first person (narrator) can tell the rumour to any one of the available $21 - 1 = 20$ persons; the second person can tell the rumour to any one of the remaining $20 - 1 = 19$ persons; the third person can tell the rumour to anyone of the remaining $20 - 2 = 18$ persons; …; the 10th person can tell the rumour to any one of the remaining $20 - 9 = 11$ persons.

Hence the favourable number of cases for E_2 are $20 \times 19 \times 18 \times \ldots \times 11$.

$\therefore$ Required probability $= P(E_2) = \frac{20 \times 19 \times 18 \times \ldots \times 11}{20^{10}}$

Example 11: If 10 men, among whom are A and B, stand in a row, what is the probability that there will be exactly 3 men between A and B?

Solution: If 10 men stand in a row, then A can occupy any one of the 10 positions and B can occupy any one of the remaining 9 positions. Hence, the exhaustive number of cases for the positions of two men A and B are $10 \times 9 = 90$.

The cases favourable to the event that there are exactly 3 men between A and B are given below:

(i) A is in the 1st position and B is in the 5th position.
(ii) A is in the 2nd position and B is in the 6th position.

. . . .

. . . .

. . . .

(vi) A is in the 6th position and B is in the 10th position.

Further, since A and B can interchange their positions, the total number of favourable cases $= 2 \times 6 = 12$.

$\therefore$ Required probability $= \frac{12}{90} = \frac{2}{15} = 0.1333$

Example12: A five digit number is formed by the digits 0, 1, 2, 3, 4 (without repetition). Find the probability that the number formed is divisible by 4.

Solution: The total number of ways in which the five digits 0, 1, 2, 3, 4 can be arranged among themselves is $5!$. Out of these, the number of arrangements which begin with 0 (and therefore will give only 4-digited numbers) is $4!$.

Hence the total number of five digited numbers that can be formed from digits 0, 1, 2, 3, 4 is $5! - 4! = 120 - 24 = 96$

The number formed will be divisible by 4 if the number formed by the two digits on extreme right (i.e., the digits in the unit and tens places) is divisible by 4. Such numbers are:

$$04, \quad 12, \quad 20, \quad 24, \quad 32 \text{ and } 40$$

If the numbers end in 04, the remaining three digits viz., 1, 2 and 3 can be arranged among themselves in 3! in each case.

If the numbers end with 12, the remaining three digits 0, 2, 3 can be arranged in $3!$ ways. Out of these we shall reject those numbers which start with 0 (i.e., have 0 as the first digit). There are $(3-1)! = 2!$ such cases. Hence, the number of five digited numbers ending with 12 is : $3! - 2! = 6 - 2 = 4$

Similarly the number of 5 digited numbers ending with 24 and 32 each is 4. Hence the total number of favourable cases is: $3 \times 3! + 3 \times 4 = 18 + 12 = 30$

Hence, required probability $= \frac{30}{96} = \frac{5}{16}$

Example13: There are four hotels in a certain town. If 3 men check into hotels in a day, what is the probability that each checks into a different hotel?

Solution: Since each man can check into any one of the four hotels in ${}^4C_1 = 4$ ways, the 3 men can check into 4 hotels in $4 \times 4 \times 4 = 64$ ways, which gives the exhaustive number of cases.

If three men are to check into different hotels, then first man can check into any one of the 4 hotels in ${}^4C_1 = 4$ ways; the second man can check into any one of the remaining 3 hotels in ${}^3C_1 = 3$ ways; and the third man can check into any one of the remaining two hotels in ${}^2C_1 = 2$ ways. Hence, favourable number of cases for each man checking into a different hotel is: ${}^4C_1 \times {}^3C_1 \times {}^2C_1 = 4 \times 3 \times 2 = 24$

$\therefore$ Required probability $= \frac{24}{64} = \frac{3}{8} = 0.375$

1.4

Theorems in Probability

In this module, we shall prove some theorems which help us to evaluate the probabilities of some complicated events in a rather simple way. In proving these theorems, we shall follow the axiomatic approach based on the three axioms given in axiomatic definition of probability in module 1.3 on definitions of probability.

In a problem on probability, we are required to evaluate probability of certain statements. These statements can be expressed in terms of set notation and whose probabilities can be evaluated using theorems in probability. Let A and B be two events in S. Certain statements in set notation are given in the following table.

S. No.	Statement	Set notation
1.	At least one of the events A or B occurs	$A \cup B$
2.	Both the events A and B occur	$A \cap B$
3.	Neither A nor B occurs	$\bar{A} \cap \bar{B}$
4.	Event A occurs and B does not occur	$A \cap \bar{B}$
5.	Exactly one of the events A or B occurs	$(\bar{A} \cap B) \cup (A \cap \bar{B}) = A \Delta B$
6.	Not more than one of the events A or B occurs	$(A \cap \bar{B}) \cup (\bar{A} \cap B) \cup (\bar{A} \cap \bar{B})$
7.	If event A occurs, so does B	$A \subset B$
8.	Events A and B are mutually exclusive	$A \cap B = \phi$
9.	Complement of event A	$\bar{A}$
10.	Sample space	S

Example 1: Let A, B and C are three events in S. Find expression for the events in set notation.

(i) only A occurs **(ii) both A and B, but not C, occur**
(iii) all three events occur **(iv) at least one occurs**
(v) at least two occur **(vi) one and no more occurs**
(Vii) two and no more occur **(viii) none occurs**

Solution:

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

(vi) $(A \cap \bar{B} \cap \bar{C}) \cup \left(\bar{A} \cap B \cap \overline{C}\right) \cup (\bar{A} \cap \bar{B} \cap C)$

(vii) $(A \cap B \cap \bar{C}) \cup (\bar{A} \cap B \cap C) \cup (A \cap \bar{B} \cap C)$

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

Theorems on Probability

Theorem 1: Probability of the impossible event is zero, i.e., $P(\phi) = 0$.

Proof: We know that $S \cup \phi = S \Rightarrow P\,(S) = P(S \cup \phi)$

$$\Rightarrow P(S) = P(S) + P\,(\phi) \text{ (Axiom 3)}$$

$$\Rightarrow P(\phi) = 0$$

Theorem 2: Probability of the complementary event $\bar{A}$ of A is given by $P(\bar{A}) = 1 - P(A)$.

Proof: Since A and $\bar{A}$ are mutually exclusive events in S,

$$A \cup \bar{A} = S \Rightarrow P(A \cup \bar{A}) = P(S) \Rightarrow P(A) + P(\bar{A}) = 1 \text{ (Axioms 2 and 3)}$$

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

Corollary 1: $\mathbf{0 \leq P(A) \leq 1}$

Proof: We have $P(A) = 1 - P(\bar{A}) \leq 1$ ($\because$ $P(\bar{A}) \geq 0$, by Axiom 1)

Further, $P(A) \geq 0$ (by Axiom 1). Therefore, $0 \leq P(A) \leq 1$

Corollary 2: $\boldsymbol{P(\phi) = 0}$

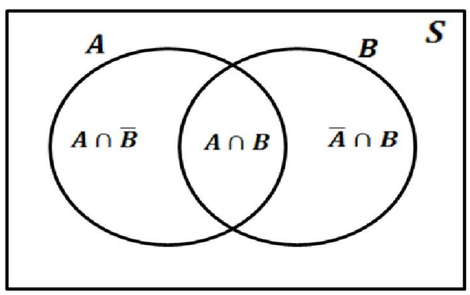
Proof: Since $\phi = \bar{S}$, $P(\phi) = P(\bar{S}) = 1 - P(S) = 1 - 1 = 0$ (by Axiom 2)

$$\Rightarrow P(\phi) = 0$$

Theorem 3: For any two events $\boldsymbol{A}$ and $\boldsymbol{B}$, we have

(i) $\boldsymbol{P(\bar{A} \cap B) = P(B) - P(A \cap B)}$ **(ii)** $\boldsymbol{P(A \cap \bar{B}) = P(A) - P(A \cap B)}$

Proof:



(i) From the Venn diagram, we have,

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

where $(\bar{A} \cap B)$ and $(A \cap B)$ are mutually exclusive events. Hence by Axiom 3,

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

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

(ii) Similarly, we have,

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

where $(A \cap B)$ and $(A \cap \bar{B})$ are mutually exclusive events. Hence by Axiom 3

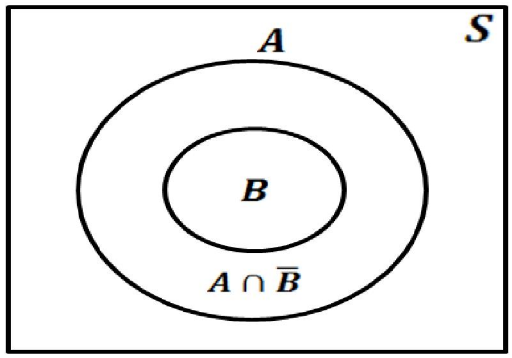
$$P(A) = P(A \cap B) + P(A \cap \bar{B})$$

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

Theorem 4: If $\boldsymbol{B \subset A}$, then

(i) $\boldsymbol{P(A \cap \bar{B}) = P(A) - P(B)}$ **(ii)** $\boldsymbol{P(B) \leq P(A)}$

Proof:



(i) If $B \subset A$, then B and $A \cap \bar{B}$ are mutually exclusive events and

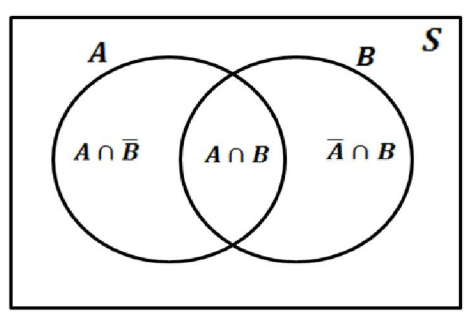
$$A = B \cup (A \cap \bar{B})$$

$$\Longrightarrow P(A) = P(B) + P(A \cap \bar{B}) \text{ (Axiom 3)}$$

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

(ii) We have $P(A \cap \bar{B}) \geq 0$ (Axiom 1). Hence $P(A) - P(B) \geq 0 \Rightarrow P(B) \leq P(A)$.
Thus, $B \subset A \Rightarrow P(B) \leq P(A)$.

Theorem 5: Addition Theorem of Probability for Two Events:
Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be any two events in $\boldsymbol{S}$. Then
$\boldsymbol{P(A \cup B) = P(A) + P(B) - P(A \cap B)}$



Proof: From Venn diagram, we have

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

where A and $\bar{A} \cap B$ are mutually exclusive events in S.

$\therefore\ P(A \cup B) = P(A) + P(\bar{A} \cap B)$ (Axiom 3)

$= P(A) + P(B) - P(A \cap B)$ (From Theorem 3)

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

Note:

1. If A and B are mutually exclusive events then $A \cap B = \phi$ and hence $P(A \cap B) = P(\phi) = 0$. Thus, if A and B are mutually exclusive events, then $P(A \cup B) = P(A) + P(B)$.
2. The addition theorem of probability for three events is given by

$$\boldsymbol{P(A \cup B \cup C) =}$$

$$\boldsymbol{P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)}$$

This can be proved first by taking $A \cup B$ as one event and C as second event and repeated application of Theorem 5

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

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

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

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

3. Addition Theorem of Probability for $\boldsymbol{n}$-Events

Let $\boldsymbol{A_1, A_2, .., A_n}$ be $\boldsymbol{n}$ events in $\boldsymbol{S}$. Then

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{\substack{i=1 \\ i<j}}^{n} \sum_{j=1}^{n} P(A_i \subset A_j) + \sum_{i=1}^{n} \sum_{\substack{j=1 \\ i<j<k}}^{n} \sum_{k=1}^{n} P(A_i \subset A_j \subset A_k) - \ldots + (-1)^{n-1} P\left(\bigcap_{i=1}^{n} A_i\right)$$

Example 2: If two dice are thrown, what is the probability that the sum is (i) greater than 8, (ii) neither 7 nor 11 (iii) an even number on the first die or a total of 8?

Solution:

(i) If two dice are thrown, then $n(S) = 6^2 = 36$. Let T be the event getting the sum of the numbers greater than 8 on the two dice. Then $T = A \cup B \cup C \cup D$, where A, B, C and D repectively the events of getting the sum of $9,10,11$ and 12. Note that A, B, C and D are pair wise mutually exclusive events. Therefore

$$P(T) = P(A) + P(B) + P(C) + P(D)$$

Note that $A = \{(3,6), (4,5), (5,4), (6,3)\}$ and $P(A) = \frac{4}{36}$

$B = \{(4,6), (5,5), (6,4)\}$ and $P(B) = \frac{3}{36}$

$C = \{(5,6), (6,5)\}$ and $P(C) = \frac{2}{36}$

$D = \{(6,6)\}$ and $P(D) = \frac{1}{36}$

$\therefore P(T) = \frac{4}{36} + \frac{3}{36} + \frac{2}{36} + \frac{1}{36} = \frac{10}{36} = \frac{5}{18}$

(ii) Let A denote the event of getting the sum of 7 and B denote the event of getting the sum of 11.Then

$A = \{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\}$ and $P(A) = \frac{6}{36}$

$B = \{(5,6), (6,5)\}$ and $P(B) = \frac{2}{36}$

$\therefore$ Required probability $= P$ (neither 7 nor 11)

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

$= 1 - [P(A) + P(B)]$ ($\because$ A and B are mutually exclusive events)

$= 1 - \left[\frac{6}{36} + \frac{2}{36}\right] = 1 - \frac{8}{36} = 1 - \frac{2}{9} = \frac{7}{9}$

(iii) Let A be the event of getting an even number on the first die and B be the event of getting the sum of 8. Therefore,

$A = \{2,4,6\} \times \{1,2,3,4,5,6\} \Rightarrow n(A) = 3 \times 6 = 18,$

$B = \{(2,6), (3,5), (4,4), (5,3), (6,2)\} \Rightarrow n(B) = 5,$

$A \cap B = \{(2,6), (4,4), (6,2)\} \Rightarrow n(A \cap B) = 3$ and
$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{18}{36} + \frac{5}{36} - \frac{3}{36} = \frac{20}{36} = \frac{5}{9}$

Example 3: A card is drawn from a pack of 52 cards. Find the probability of getting a king or a heart or a red card.

Solution: Let us define the following events:

A : The card drawn is a king

B : The card drawn is a heart

C : The card drawn is a red card

Then, A, B and C are not mutually exclusive.

$n(A) = 4$, $n(B) = 13$, $n(C) = 26$, $n(A \cap B) = 1$, $n(A \cap C) = 2$,
$n(B \cap C) = 13$, $n(A \cap B \cap C) = 1$.

$$P(A \cup B \cup C)$$

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

$$= \frac{4}{52} + \frac{13}{52} + \frac{26}{52} - \frac{1}{52} - \frac{2}{52} - \frac{13}{52} + \frac{1}{52} = \frac{28}{52} = \frac{7}{13}$$

Compound event: The simultaneous occurrence of two or more events is termed as compound event.

Compound probability: The probability of a compound event is known as compound probability.

Conditional probability: The probability of an event A occuring when it is known that some event B has occurred, is called a conditional probability of the event A, given that B has occurred and denoted by $\boldsymbol{P(A|\ B)}$.

Definition: The conditional probability of the event A, given that B has occurred, denoted by $P(A|B)$, is defined by

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \text{ if } P(B) > 0$$

If $P(B) = 0$, $P(A|B)$ is not defined.

Example 4: Consider a family with two children. Assume that each child is likely to be a boy as it is to be a girl. What is the conditional probability that both children are boys, given that (i) the older child is a boy (ii) at least one of the child is a boy?

Solution: We have the sample space $S = \{(b,b),(b,g),(gg\ ),(gg\ )\}$.Define the events:

A : Older child is a boy

B : Younger child is a boy

Therefore, $A = \{(b,b),(b,g)\}$, $P(A) = \frac{n(A)}{n(S)} = \frac{2}{4} = \frac{1}{2}$, $B = \{(b,b),(gb\ )\}$

Then $A \cap B$: both are boys, $A \cap B = \{(b,b)\}$ and $P(A \cap B) = \frac{n(A \cap B)}{n(s)} = \frac{1}{4}$

$A \cup B$: At least one is a boy

and $P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{1}{2} + \frac{1}{2} - \frac{1}{4} = \frac{3}{4}$

(i) $P\big((A \cap B)|A\big) = \frac{P(A \cap B)}{P(A)} = \frac{\frac{1}{4}}{\frac{1}{2}} = \frac{1}{2}$

(ii) $P\big((A \cap B)|(A \cup B)\big) = \frac{P[(A \cap B) \cap (A \cup B)]}{P(A \cup B)} = \frac{P(A \cap B)}{P(A \cup B)} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}$

Independent events: Two events A and B are said to be independent if the happening or non-happening of A is not affected by the happening or non-happening of B. Thus, A and B are independent if and only if the conditional probability of the event A given that B has happened is equal to the probability of A. That is,

$$P(A|B) = P(A) \text{ if } P(B) > 0$$

Similarly $P(B|A) = P(B)$ if $P(A) > 0$

By the definition of conditional probability, we have

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

Thus, A and B are independent events, if and only if

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

In general, $A_1, A_2, \dots, A_n$ are independent events, if and only if

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1) \cdot P(A_2) \cdot \dots \cdot P(A_n)$$

Pair wise Independent Events: A set of events $A_1, A_2, \dots, A_n$ are said to be pairwise independent if every pair of different events are independent.

That is, $P(A_i \cap A_j) = P(A_i) \cdot P(A_j)$ for all i and j, $i \neq j$.

Mutual Independent Events: A set of events $A_1, A_2, \dots, A_n$ are said to be mutually independent, if $P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \cdot P(A_{i_2}) \cdot \dots \cdot P(A_{i_k})$ for every subset $\{A_{i_1}, A_{i_2}, \dots, A_{i_k}\}$ of $\{A_1, A_2, \dots, A_n\}$.

Note: Pair wise independence does not imply mutual independence.

Theorem 6: Multiplication Theorem for Two events

Let A and B be any two events, then

$$P(A \cap B) = \begin{cases} P(A).P(B|A) & \text{if } P(A) > 0 \\ P(B).P(A|B) & \text{if } P(B) > 0 \\ P(A).P(B) & \text{if A and B are independent} \end{cases}$$

The proof follows from definition of conditional probability.

Note: Multiplication Theorem for n-Events $A_1, A_2, A_3, \ldots, A_n$

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = \begin{cases} P(A_1).P(A_2 \mid A_1).P(A_3 \mid (A_1 \cap A_2))\ldots P(A_n \mid (A_1 \cap A_2 \cap \ldots \cap A_{n-1})) \\ P(A_1).P(A_2).P(A_3)\ldots P(A_n), \text{if } A_1, A_2, \ldots, A_n \text{ are independent} \end{cases}$$

Theorem 7: If A_1 and A_2 are independent events, then A_1 and $\overline{A_2}$ are also independent.

Proof: **(See P3)**

Theorem 8: If A_1 and A_2 are independent events, then $\overline{A_1}$ and $\overline{A_2}$ are also independent.

Proof: **(See P4)**

Example 5: A fair dice is thrown twice. Let A, B and C denote the following events:

A : First toss is odd; B : Second toss is even; C Sum of numbers is 7

- **(i) Find $P(A), P(B)$ and $P(C)$.**
- **(ii) Show that A, B and C are pair wise independent**
- **(iii) Show that A, B and C are not independent**

Solution:

(i) The number of outcomes in the sample space S is given by $n(S) = 6^2 = 36$.
We have,

$$A = \{1,3,5\} \times \{1,2,3,4,5,6\} \text{ and } n(A) = 3 \times 6 = 18$$

$$B = \{1,2,3,4,5,6\} \times \{2,4,6\} \text{ and } n(B) = 6 \times 3 = 18$$

$$C = \{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\} \text{ and } n(C) = 6$$

Therefore, $P(A) = \frac{18}{36} = \frac{1}{2}$, $P(B) = \frac{18}{36} = \frac{1}{2}$ and $P(C) = \frac{6}{36} = \frac{1}{6}$.

(ii) $A \cap B = \{(1,2),(1,4),(1,6),(3,2),(3,4),(3,6),(5,2),(5,4),(5,6)\}$

$\therefore n(A \cap B) = 9$ and $P(A \cap B) = \frac{9}{36} = \frac{1}{4}$

But $P(A) \cdot P(B) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$

Thus, $P(A \cap B) = P(A) \cdot P(B) \Rightarrow A$ and B are independent.

Next consider $A \cap C = \{(1,6),(3,4),(5,2)\}$

$\therefore n(A \cap C) = 3$ and $P(A \cap C) = \frac{3}{36} = \frac{1}{12}$.

But $P(B) \cdot P(C) = \frac{1}{2} \times \frac{1}{6} = \frac{1}{12}$.

Thus, $P(B \cap C) = P(B) \cdot P(C) \Rightarrow B$ and C are independent

(iii) Consider $A \cap B \cap C = \{(1,6),(3,4),(5,2)\}$

$\therefore n(A \cap B \cap C) = 3$ and $P(A \cap B \cap C) = \frac{3}{36} = \frac{1}{12}$

But $P(A) \cdot P(B) \cdot P(C) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{6} = \frac{1}{24}$

Thus, $P(A \cap B \cap C) \neq P(A) \cdot P(B) \cdot P(C)$

$\Rightarrow A, B$ and C are not independent.

Theorem 9: If A_1 and A_2 are independent events, then

$$\boldsymbol{P(A_1 \cup A_2) = 1 - P(\overline{A_1}) \cdot P(\overline{A_2})}$$

Proof: Consider $RHS = 1 - P(\overline{A_1}) \cdot P(\overline{A_2})$

$$= 1 - \left[\left(1 - P(A_1)\right) \cdot \left(1 - P(A_2)\right) \right]$$

$$= 1 - \left(1 - P(A_1) - P(A_2) + P(A_1) \cdot P(A_2)\right)$$

$$= 1 - 1 + P(A_1) + P(A_2) - P(A_1) \cdot P(A_2)$$

$$= P(A_1) + P(A_2) - P(A_1 \cap A_2)$$

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

Thus, $P(A_1 \cup A_2) = 1 - P(\overline{A_1}) \cdot P(\overline{A_2})$.

Generalization: If $\boldsymbol{A_1, A_2, \dots, A_n}$ are $\boldsymbol{n}$ independent events, then

$$\boldsymbol{P(A_1 \cup A_2 \cup \dots \cup A_n) = 1 - P(\overline{A_1}) \cdot P(\overline{A_2}) \cdot \dots \cdot P(\overline{A_n})}$$

Example 6: A problem in probability is given to three students $\boldsymbol{A, B}$ and $\boldsymbol{C}$ whose chances of solving it are $\boldsymbol{\frac{1}{3}}$, $\boldsymbol{\frac{1}{4}}$ and $\boldsymbol{\frac{1}{5}}$ respectively. Find the probability that the problem will be solved if they all try independently.

Solution: Let E_1, E_2 and E_3 denote the events that the problem is solved by A, B and C respectively. Then, we have

$$P(E_1) = \frac{1}{3} \Rightarrow P(\overline{E_1}) = \frac{2}{3}$$

$$P(E_2) = \frac{1}{4} \Rightarrow P(\overline{E_2}) = \frac{3}{4}$$

$$P(E_3) = \frac{1}{5} \Rightarrow P(\overline{E_3}) = \frac{4}{5}$$

The problem is solved if atleast one of them is able to solve it.

Thus, $P(E_1 \cup E_2 \cup A_3) = 1 - P(\overline{E_1}) \cdot P(\overline{E_2}) \cdot P(\overline{E_3}) = 1 - \frac{2}{3} \times \frac{3}{4} \times \frac{4}{5} = 1 - \frac{2}{5} = \frac{3}{5}$

1.5

Bayes‘Theorem and Its Applications

One of the important applications of the conditional probability is in the computation of unknown probabilities on the basis of the information supplied by the experiment or past records. For example, suppose an event has occurred through one of the various mutually exclusive events or reasons. Then the conditional probability that it has occurred due to a particular event or reason is called it as **inverse** or **posteriori probability**. These probabilities are computed by Bayes' theorem, named so after the British mathematician **Thomas Bayes** who propounded it in 1763. The revision of old (given) probabilities in the list of the additional information supplied by the experiment or past records is of extreme help in arriving at valid decisions in the face of uncertainty.

Bayes' Theorem (Rule for the Inverse Probability)

Let $E_1, E_2, \dots, E_n$ be n be mutually exclusive and exhaustive events in the sample space S with $P(E_i) \neq 0$ for $i = 1, 2, \dots, n$. Let A be an arbitrary event which is a subset of $\bigcup_{i=1}^{n} E_i$ such that $P(A) > 0$. Then

$$P(E_i \mid A) = \frac{P(E_i).P(A|E_i)}{\sum_{i=1}^{n} P(E_i).P(A|E_i)} = \frac{P(E_i).P(A|E_i)}{P(E_i)} \qquad \text{for } i = 1, 2, \dots, n$$

Proof: Since $A \subset \bigcup_{i=1}^{n} E_i$, we have $A = A \cap \left(\bigcup_{i=1}^{n} E_i \right) = \bigcup_{i=1}^{n} (A \cap E_i)$.

Since $(A \cap E_i) \subset E_i (i = 1, 2, \dots, n)$ are mutually exclusive events, we have by addition theorem of probability

$$P(A) = P\left(\bigcup_{i=1}^{n} (A \cap E_i) \right)$$

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

$$\Rightarrow P(A)=\sum_{i=1}^{n} P(E_i).P(A|E_i)$$ (By multiplication theorem of probability)

Also we have

$$P(A \cap E_i) = P(E_i|A).P(A)$$

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

$$\Rightarrow P(E_i|A)=\frac{P(E_i).P(A|E_i)}{\sum_{i=1}^{n} P(E_i).P(A|E_i)} \quad \text{for } i = 1, 2, \dots, n$$

which is the Bayes' rule.

Note:

1. The probabilities $P(E_1), P(E_2), \dots, P(E_n)$ are known as the 'a priori probabilities', because they exist before we gain any information from the experiment itself.
2. The probabilities $P(A|\, E_i)$ $i = 1, 2, \dots n$ are called 'likelihoods' because they indicate how likely the event A under consideration is to occur, given each and every a priori probability.
3. The probabilities $P(E_i|A), i = 1, 2, .., n$ are called 'posteriori probabilities' because they are determined after the results of the experiment are known.
4. $P(A)=\sum_{i=1}^{n} P(E_i).P(A|E_i)$ is known as **total probability**.
5. Bayes' theorem is extensively used by *business, management* and *engineering* executives in arriving at valid decisions in the face of uncertainty.

Example 1: In a bolt factory machines A, B, C manufacutre respectively 25%, 35% and 40% of the total. Of their output $5, 4, 2$ percent are known to be defective bolts. A bolt is drawn at random from the product and is found to be defective. What are the probabilities that it was manufactured by

(i) Machine A.

(ii) Machine B or C

Solution: Let E_1, E_2 and E_3 denote respectively the events that the bolt selected at random is manufactured by the machines A, B and C respectively and let E denote the event that it is defective. Then we have:

E_i	E_1	E_2	E_3	Total
$P(E_i)$	0.25	0.35	0.40	1
$P(E\|E_i)$	0.05	0.04	0.02	
$P(E \cap E_i) = P(E_i).P(E\|E_i)$	0.0125	0.0140	0.0080	$P(E) = 0.0345$

$$P(E) = \sum_{i=1}^{3} P(E_i).P(E|E_i) = 0.0345$$

(i) Hence, the probability that a defective bolt chosen at random is manufactured by factory A is given by Bayes' rule as:

$$P(E_1|E) = \frac{P(E_1)P(E|E_1)}{\sum P(E_i)P(E|E_i)} = \frac{0.0125}{0.0345} = 0.36$$

(ii) Similarly,

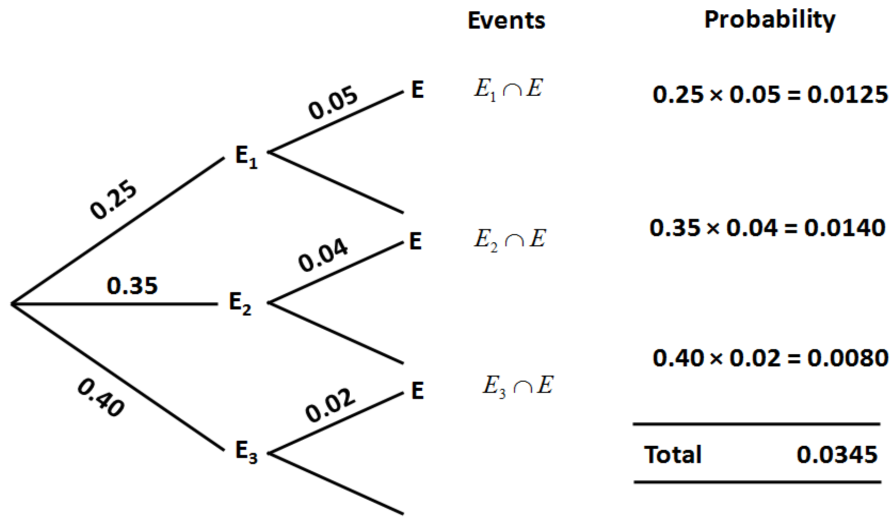
$$P(E_2|E) = \frac{0.0140}{0.0345} = \frac{28}{69} = 0.41\,; \qquad P(E_3|E) = \frac{0.0080}{0.0345} = \frac{16}{69} = 0.23$$

Hence, the probability that a defective bolt chosen at random is manufactured by machine B or C is:

$$P(E_2|E) + P(E_3|E) = 0.41 + 0.23 = 0.64$$

(OR Required probability is equal to $1 - P(E_1|E) = 1 - 0.36 = 0.64$)

Aliter: **TREE DIAGRAM**



From the above diagram the probability that a defective bolt is manufactured by factory A is

$$P(E_1|E) = \frac{0.0125}{0.0345} = 0.36$$

Similarly, $P(E_2|E) = \frac{0.0140}{0.0345} = 0.41$ and $P(E_3|E) = \frac{0.0080}{0.0345} = 0.23$

Hence, the probability that a defective bolt chosen at random is manufactured by machine B or C is:

$$P(E_2|E) + P(E_3|E) = 0.41 + 0.23 = 0.64$$

(OR Required probability is equal to $1 - P(E_1|E) = 1 - 0.36 = 0.64$)

Remark: Since $P(E_3)$ is greatest, on the basis of 'a *priori*' probabilities alone, we are likely to conclude that a defective bolt drawn at random from the product is manufactured by machine C. After using the additional information we obtained the *posterior* probabilities which give $P(E_2|E)$ as maximum. Thus, we shall now say that it is probable that the defective bolt has been manufactured by machine B, a result which is different from the earlier conclusion. However, latter conclusion is a much valid conclusion as it is based on the entire information at our disposal. Thus, Bayes's rule provides a very powerful tool in improving the quality of probability and this helps the management executives in arriving at valid decisions in the face of uncertainty. Thus, the additional information reduces the importance of the prior probabilities. The only requirement for the use of *Bayesian rule* is that all the hypotheses under consideration must be valid and that none is assigned 'a prior' probability 0 or 1.

Example 2: In a railway reservation office, two clerks are engaged in checking reservation forms. On an average, the first clerk checks 55% of the forms, while the second does the remaining. The first clerk has an error rate of 0.03 and second has an error rate of 0.02 , A reservation form is selected at random from the total number of forms checked during a day, and is found to have an error. Find the probability that it was checked (i) by the first (ii) by the second clerk.

Solution: Let us define the following events:

E_1: The selected form is checked by clerk 1.

E_2: The selected form is checked by clerk 2.

E: The selected form has an error.

Then we are given:

$$P(E_1) = 55\% = 0.55 \ ; \quad P(E_2) = 45\% = 0.45 \ ;$$

$$P(E|E_1) = 0.03 \quad ; \quad P(E|E_2) = 0.02$$

Required to find $P(E_1|E)$ and $P(E_2|E)$. By Bayes' Rule the probability that the form containing the error was checked by clerk 1, is given by;

$$P(E_1|E) = \frac{P(E_1)\ P(E_1|E)}{P(E_1)\ P(E|E_1)+P(E_2)P(E|E_2)} = \frac{0.55\times0.03}{0.55\times0.03+0.45\times0.02}$$

$$= \frac{0.0165}{0.0165+0.0090} = \frac{0.0165}{0.0255} = 0.647$$

Similarly, the probability that the form containing the error was checked by clerk 2, is given by

$$P(E_2|E) = \frac{P(E_2)\ P(E|E_2)}{P(E_1)P(E|E_1)+P(E_2)P(E|E_2)} = \frac{0.45\times0.02}{0.55\times0.03+0.45\times0.02} = \frac{0.0090}{0.0255} = 0.353$$

(OR $P(E_2|E) = 1 - P(E_1|E) = 1 - 0.647 = 0.353$)

Example 3: The results of an investigating by an expert on a fire accident in a skyscraper are summarized below:

(i) **Prob. (there could have been short circuit)** $= \mathbf{0.8}$

(ii) **Prob. (LPG cylinder explosion)** $= \mathbf{0.2}$

(iii) **Chance of fire accident is** 30% **given a short circuit and** 95% **given an LPG explosion.**

Based on these, what do you think is the most probable cause of fire?

Solution: Let us define the following events:

E_1: Short circuit ; E_2: LPG explosion ; E: Fire accident

Then, we are given:

$$P(E_1) = 0.8 \;; \quad P(E_2) = 0.2 \;;$$

$$P(E|E_1) = 0.30 \;; \quad P(E|E_2) = 0.95$$

By Bayes' Rule:

$$P(E_1|E) = \frac{P(E_1).P(E|E_1)}{P(E_1)P(E|E_1)+P(E_2)P(E|E_2)} = \frac{0.80\times 0.30}{0.8\times 0.30+0.2\times 0.95} = \frac{0.240}{0.240+0.190} = \frac{24}{43}$$

$$P(E_2|E) = \frac{P(E_2)P(E|E_2)}{P(E_1)P(E|E_1)+P(E_2)P(E|E_2)} = \frac{0.190}{0.430} = \frac{19}{43}$$

(OR $P(E_2|E) = 1 - P(E_1|E) = 1 - \frac{24}{43} = \frac{19}{43}$)

Since $P(E_1|E) > P(E_2|E)$, short circuit is the most probable cause of fire.

Example 4: The contents of urns I, II and III are respectively as follows:

1 white, 2 black and 3 red balls,

2 white, 1 black and 1 red balls, and

4 white, 5 black and 3 red balls.

One urn is chosen at random and two balls drawn. They happen to be white and red. What is the probability that they came from urns I, II, III?

Solution:

Let E_1, E_2 and E_3 denote the events of choosing 1st , 2nd and 3rd urn respectively and let E be the event that the two balls drawn from the selected urn are white and red. Then we have:

	E_1	E_2	E_3
$P(E_i)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
$P(E\|E_i)$	$\frac{1\times 3}{6_{C_2}}=\frac{1}{5}$	$\frac{2\times 1}{4_{C_2}}=\frac{1}{3}$	$\frac{4\times 3}{12_{C_2}}=\frac{2}{11}$
$P(E\cap E_i)=P(E_i)\times P(E\|E_i)$	$\frac{1}{3}\times\frac{1}{5}=\frac{1}{15}$	$\frac{1}{3}\times\frac{1}{3}=\frac{1}{9}$	$\frac{1}{3}\times\frac{2}{11}=\frac{2}{33}$

We have:

$$\sum P(E_i)\,P(E|E_i)=\frac{1}{15}+\frac{1}{9}+\frac{2}{33}=\frac{33+55+30}{495}=\frac{118}{495}$$

Hence by Bayes's rule, the probability that the two white and red balls drawn are from 1st urn is:

$$P(E_1|E)=\frac{P(E_1)P(E|E_1)}{\sum P(E_i)P(E|E_i)}=\frac{\frac{1}{15}}{\frac{118}{495}}=\frac{33}{118}$$

Similarly, we have

$$P(E_2|E) = \frac{P(E_2)P(E|E_2)}{\sum P(E_i)P(E|E_i)} = \frac{\frac{1}{9}}{\frac{118}{495}} = \frac{55}{118}$$

and $\quad P(E_3|E) = \frac{\frac{2}{33}}{\frac{118}{495}} = \frac{30}{118}$ $\qquad$ (Or $\quad P(E_3|E) = 1 - \frac{33}{118} - \frac{55}{118} = \frac{30}{118}$)

Unit – 2

Probability distributions

2.1

Random Variable

While performing a random experiment we are mainly concerned with the assignment and computation of probabilities of events. In many experiments we are interested in some function of the outcomes of the experiment as opposed to the outcome itself. For instance, in tossing two dice we are interested in the sum of faces of the dice and are not really concerned about the actual outcome. That is, we may be interested in knowing that the sum is seven and not be concerned over whether actual outcome was $(1,6)$ or $(2,5)$ or $(3,4)$ or $(4,3)$ or $(5,2)$ or $(6,1)$. These quantities of interest or more formally these real valued function defined on the sample space are known as **random variables**.

Random variable (r. v): Let S be the sample space associated with a random experiment. Let $\boldsymbol{R}$ be the set of real numbers. If $X: S \longrightarrow \boldsymbol{R}$, $i.e.$, X is a real valued function defined on the sample space, then X is known as a **random variable**. In other words, random variable is a function which takes real values which are determined by the outcomes in the sample space.
The random variables are denoted by capital letters X, Y, Z … etc.

Notation: Let $a, b \in \boldsymbol{R}$. The set of all ω in S such that $X(\omega) = a$ is denoted by $X = a$. That is, $X = a$ denotes the event $\{\omega \in S | X(\omega) = a\}$. Similarly $X \leq a$ denotes the event $\{\omega \in S | X(\omega) \leq a\}$ and $a < X \leq b$ denotes the event $\{\omega \in S | X(\omega) \in (a, b]\}$.

Let us consider a random experiment of three tosses of a coin. Then the sample space S consits of $2^3 = 8$ points as given below.

$$S = \{H, T\} \times \{H, T\} \times \{H, T\}$$

$$= \{HH, HT, TH, TT\} \times \{H, T\}$$

$$= \{HHH, HTH, THH, TTH, HHT, HTT, THT, TTT\}$$

For each outcome ω in S define $X(\omega)$ as the number of heads in the outcome ω. Then X may take any one of the values $0, 1, 2$ or 3. For each outcome in S, we have one value of X. Thus,

$$X(HHH) = 3\ ,$$

$$X(HTH) = X(THH) = X(HHT) = 2$$

$$X(TTH) = X(HTT) = X(THT) = 1 \text{ and}$$

$$X(TTT) = 0$$

This shows that X is a random variable.

Note that $X = 0, X = 1, X = 2$ and $X = 3$ respectively denote the events

$\{TTT\}, \{TTH, HTT, THT\}, \{HTH, THH, HHT\}$ and $\{HHH\}$

Discrete Random Variable (d. r. v): If the random variable assumes only a finite or countably infinite set of values, it is known as **discrete random variable**.
For example, the number of students attending the class, the number of defectives in a lot consisting of manufactured items and the number of accidents taking place on a busy road, etc., are all discrete random variables. In the above example X is a d.r.v.

Continuous Random Variable (c. r. v): If a random variable can assume uncountable set of values, it is said to a **continuous random variable**.
For example, the age, height or weight of the students in a class is all continuous random variables. In case of continuous random variable, we usually talk of the value in a particular interval and not at a point. Generally, discrete random

variable represents *count data* while continuous random variable represent *measured data*.

The probabilistic behavior of a d.r.v.X at each real point is described by a function called **probability mass function** and it is defined below:

Probability Mass Function (p.m.f): Let X be a discrete random variable with distinct values $x_1, x_2, \ldots, x_n, \ldots$ .The function $p : \boldsymbol{R} \rightarrow \boldsymbol{R}$ defined as

$$p(x) = \begin{cases} P(X = x_i) = p_i & if \;\; x = x_i \\ 0 & if \;\; x \neq x_i, i = 1,2, \ldots, n, \ldots \end{cases}$$

is called the **probability mass function** of r.v.X, if (i) $p(x) \geq 0 \; \forall x \in \boldsymbol{R}$ and (ii) $\sum_{x \in R} p(x) = 1$

Probability Distribution: The set of all possible ordered pairs $(x_i, p(x_i)), i = 1,2, \ldots, n, \ldots$ is called the **probability distribution** of the r.v.X.

In particular, if X takes the values $x_1, x_2, x_3, \ldots, x_n$ then the probability of X is usually represented in a tabular form as given below:

Probability Distribution of r. v. x

x	x_1	x_2	x_3	…	x_n
$p(x)$	p_1	p_2	p_3	…	p_n

Note: The concept of probability distribution is analogous to that of frequency distribution. Just as frequency distribution tells us how the total frequency is distributed among different values (or classes) of the variable, similarly a probability distribution tells us how total probability 1 is distributed among the various values which the r. v. can take.

Example 1: Obtain the probability distribution of X, the number of heads in three tosses of a coin (or a simultaneous toss of three coins).

Solution:

The sample space S consists of $2^3 = 8$ sample points, as give below:

$$S = \{H,T\} \times \{H,T\} \times \{H,T\} = \{HH, HT, TH, TT\} \times \{H,T\}$$

$$= \{HHH, HTH, THH, TTH, HHT, HTT, THT, TTT\}$$

Obviously, X is a random variable which can take the values 0, 1, 2 or 3.

The probability distribution of X is computed as given below.

No. of heads $\boldsymbol{a}$	$\boldsymbol{X = a}$ $\{\boldsymbol{\omega \in S \mid X(\omega) = a}\}$	**No. of favourable cases**	$\boldsymbol{p(x) = P(X = a)}$
0	$\{TTT\}$	1	$\frac{1}{8}$
1	$\{TTH, HTT, THT\}$	3	$\frac{3}{8}$
2	$\{HTH, THH, HHT\}$	3	$\frac{3}{8}$
3	$\{HHH\}$	1	$\frac{1}{8}$

Hence, the probability distribution of X is given by:

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

Probability Density Function (p. d. f): Let X be a continuous random variable defined on the sample space S. Let $f(x)$ be a real valued function defined on $\boldsymbol{R}$ such that, for any real numbers a and $b (a < b)$, $P(a \le X \le b) = \int_a^b f(x)dx.$

If the function $f(x)$ satisfies (i) $f(x) \ge 0 \ \ \forall x \in \boldsymbol{R}$ and (ii) $\int_{-\infty}^{\infty} f(x)dx = 1$ then $f(x)$ is known as probability density function (p.d.f) of X

Note:

1. If X is a c. r. v., then $P(X = a) = 0$ where a is some real number.
2. Unlike discrete probability distribution, a continuous probability distribution can't be presented in a tabular form.

Cumulative Distribution Function (c. d. f): The cumulative distribution function of a r. v. X is defined by

$$F(x) = P(X \leq x) = = \begin{cases} \sum_{t \leq x} p(t) & \text{if } X \text{ is a d.r.v. with p.m.f } p(x) \\ \int_{-\infty}^{x} f(t)dt & \text{if } X \text{ is a c.r.v. with p.d.f } f(x) \end{cases}$$

Note: If X is a continuous random variable, then $\frac{d}{dx}F(x) = f(x)$

Properties of c.d.f.

1. If $a < b, P(a < X \leq b) = F(b) - F(a)$
2. $0 \leq F(x) \leq 1$ and $F(x) \leq F(y)$ if $x < y$
3. $F(-\infty) = 0$ and $F(\infty) = 1$
4. Discontinuities of $F(x)$ are atmost countable.

Note: The c.d.f. is used to find the cumulative probabilities in a probability distribution.

Example 2:

(i) Find the constant k such that

$$f(x) = \begin{cases} kx^2 & , \quad 0 < x < 3 \\ 0 & , \quad \text{otherwise} \end{cases}$$

is a p.d.f.

(ii) Compute $P(1 < x < 2)$

(iii) Find the c.d.f and use it to compute $P(1 < x \leq 2)$

Solution:

(i) $f(x)$ is a p.d.f if

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

$$\Rightarrow k\int_0^3 x^2 dx \Rightarrow k\left[\frac{x^3}{3}\right]_0^3 = 1 \Rightarrow k = \frac{1}{9}$$

$$f(x) = \begin{cases} \frac{1}{9}x^2 & , \quad 0 < x < 3 \\ 0 & , \quad \text{otherwise} \end{cases}$$

(ii)

$$P(1 < x < 2) = \int_1^2 f(x)dx$$

$$= \int_1^2 \frac{1}{2}x^2 dx = \frac{1}{9}\left[\frac{x^3}{3}\right]_1^2 = \frac{7}{27}$$

(iii) We have,

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(u)du$$

If $x < 0$, then $F(x) = 0$. If $0 \le x < 3$, then

$$F(x) = \int_{-\infty}^{x} f(u)du = \frac{1}{9}\int_0^x u^2 du = \frac{x^3}{27}$$

If $x \ge 3$, then

$$F(x) = \int_0^3 f(u)du + \int_3^x f(u)du = \frac{1}{9}\int_0^3 u^2 du + \int_3^x du = \frac{1}{9}\times 9 + 0 = 1$$

Thus, required c.d.f is

$$F(x) = \begin{cases} 0 & , \quad x < 0 \\ \frac{x^3}{27} & , \quad 0 \le x < 3 \\ 1 & , \quad x \ge 3 \end{cases}$$

Hence $P(1 < x \le 2) = P(x \le 2) - P(x \le 1)$

$$= F(2) - F(1)$$

$$= \frac{2^3}{27} - \frac{1^3}{27} = \frac{8}{27} - \frac{1}{27} = \frac{7}{12}$$

Example 3: **A die is tossed twice. Getting *an odd number* is termed as a success. Find the probability distribution and c.d.f of the number of successes.**

Solution: Since the cases favorable to getting an odd number in a throw of a die are1,3,5, i.e., 3 in all.

Probability of success$(S) = \frac{3}{6} = \frac{1}{2}$; Probability of failure $(F) = 1 - \frac{1}{2} = \frac{1}{2}$.

If X dentoes the number of successes in two throws of a die, then X is a random variable which takes the values 0,1,2.

$$P(X = 0) = P[F \text{ in } 1^{\text{st}} \text{ throw and } F \text{ in } 2^{\text{nd}} \text{ throw}]$$

$$= P(FF) = P(F) \times P(F) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}.$$

$$P(X = 1) = P(S \text{ and } F) + P(F \text{ and } S)$$

$$= P(S)P(F) + P(F)P(S)$$

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

$$P(X = 2) = P(S \text{ and } S) = P(S)P(S) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4} \quad .$$

Hence the probability distribution of X is given by :

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

The c.d.f is given by

$$F(x) = \begin{cases} 0 & if \quad x < 0 \\ \frac{1}{4} & if \quad 0 \le x < 1 \\ \frac{3}{4} & if \quad 1 \le x < 2 \\ 1 & if \quad x \ge 2 \end{cases}$$

Example 4: Two cards are drawn

(a) successively with replacement

(b) simultaneously (successively without replacement),

from a well shuffled deck of 52 cards. Find the probability distribution of the number of aces.

Solution: Let X denote the number of aces obtained in a draw of two cards. Obviously, X is a random variable which can take the values 0, 1 or 2.

(a) Probability of drawing an ace is $\frac{4}{52} = \frac{1}{13}$

$\Longrightarrow$ Probability of drawing a non-ace is$1 - \frac{1}{13} = \frac{12}{13}$.

Since the cards are drawn with replacement, all the draws are independent.

$$P(X = 2) = P(\text{Ace and Ace}) = P(\text{Ace}) \times P(\text{Ace}) = \frac{1}{13} \times \frac{1}{13} = \frac{1}{169}$$

$$P(X = 1) = P(\text{Ace and Non-ace}) + P(\text{Non-ace and Ace})$$

$$= P(\text{Ace}) \times P(\text{Non-ace}) + P(\text{Non-ace}) \times P(\text{Ace})$$

$$= \frac{1}{13} \times \frac{12}{13} + \frac{12}{13} \times \frac{1}{13} = \frac{24}{169}.$$

$$P(X = 0) = P(\text{Non-ace and Non-ace})$$

$$= P(\text{Non-ace}) \times P(\text{Non-ace}) = \frac{12}{13} \times \frac{12}{13} = \frac{144}{169}.$$

Hence, the probability distribution of X is given by:

x :	0	1	2
$p(x)$:	$\frac{144}{169}$	$\frac{24}{169}$	$\frac{1}{169}$

(b) If cards are drawn without replacement, then exhaustive number of cases of drawing 2 cards out of 52 cards is ${}^{52}C_2$.

$\therefore P(X=0) = P(\text{No ace}) = P(\text{Both cards are non-aces})$

$$= \frac{^{48}C_2}{^{52}C_2} = \frac{48\times 47}{52\times 51} = \frac{188}{221}$$

$P(X=1) = P(\text{one ace}) = P(\text{one ace and one non-ace})$

$$= \frac{^{4}C_1 \times {}^{48}C_1}{^{52}C_2} = \frac{4\times 48\times 2}{52\times 51} = \frac{32}{221}$$

$P(X=2) = P(\text{both aces}) = \frac{^{4}C_2}{^{52}C_2} = \frac{4\times 3}{52\times 51} = \frac{1}{221}$

Hence, the probability distribution of X is given by :

x :	0	1	2
$p(x)$:	$\frac{188}{221}$	$\frac{32}{221}$	$\frac{1}{221}$

Example 5: **If X is a continuous random variable with p.d.f**

$$f(x) = \begin{cases} kx\ ,\ 0 \le x < 1 \\ k\ \ ,1 \le x < 2 \\ -k(x-3)\ \ ,2 \le x < 3 \\ 0\ \ ,\text{elsewhere} \end{cases}$$

(i) **Determine k.**

(ii) **Compute $P(x \le 1.5)$**

Solution:

(i) Since $f(x)$ is the p.d.f, so we have

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

$$\Rightarrow \int_{0}^{1} f(x)\,dx + \int_{1}^{2} f(x)\,dx + \int_{2}^{3} f(x)\,dx = 1$$

$$\Rightarrow \int_{0}^{1} kx\,dx + \int_{1}^{2} k\,dx + \int_{2}^{3} -k(x-3)\,dx = 1$$

$$\Rightarrow k\left[\frac{x^2}{2}\right]_0^1 + k[x]_1^2 - k\left[\frac{x^2}{2} - 3x\right]_2^3 = 1$$

$$\Rightarrow \frac{k}{2} + 2k - k - k\left[\left(\frac{9}{2} - 9\right) - (2-6)\right] = 1$$

$$\Rightarrow k\left[\frac{1}{2} + 2 - 1 - \frac{9}{2} + 9 + 2 - 6\right] = 1$$

$$\Rightarrow 2k = 1 \Rightarrow k = \frac{1}{2}$$

(ii)

$$P(x \le 1.5) = \int_{-\infty}^{1.5} f(x)\,dx = \int_{0}^{1} f(x)\,dx + \int_{-\infty}^{1.5} f(x)\,dx$$

$$= k\int_{0}^{1} x\,dx + \int_{1}^{1.5} k\,dx = k\left[\frac{x^2}{2}\right]_0^1 + k[x]_1^{1.5} = k\left[\frac{1}{2} + \frac{1}{2}\right] = k = \frac{1}{2}$$

2.2

Bivariate random variable

In the real life situations more than one variable effects the outcome of a random experiment. For example, consider an electronic system consisting of two components. Suppose the system will fail if both the components fail. The probability distribution of the life of the system depends jointly on the probability distributions of lives of the components. Knowing the probability distributions of lives of the components will not provide us the enough information. What we need is the probability distribution of the simultaneous behavior of lives of the components. A pair of random variables is known as a **bivariate random variable.** The individual random variables in the pair may be related.

Bivariate random variable: let S be the sample space associated with a random experiment. Let $\boldsymbol{R}$ be the real line. If (X, Y): $S \longrightarrow R \times R$, $i.e., (X, Y)(\omega) = \big(X(\omega), Y(\omega)\big)$ $\forall\ \omega \in S$, then the pair (X, Y) is known as a bivariate random variable.

Note:

1. If X and Y are both discrete random variables, then (X, Y) is a bivariate discrete random variable.
2. If X and Y are both continuous random variables, then (X, Y) is a bivariate continuous random variable.

Joint probability mass function: Let (X, Y) be a bivariate discrete random variable, which takes the values (x_i, y_j) for $i = 1,2,3, \dots, m$ and $j = 1,2,3, \dots, n$. Let

$$p(x_i, y_j) = P(X = x_i, Y = y_j)\ \forall\ i \text{ and } j$$

Then $p(x_i, y_j) \geq 0\ \forall\ i$ and j and $\sum_{i=1}^{m} \sum_{j=1}^{n} p(x_i, y_j) = 1$. The function $p(x, y)$ is known as **joint probability mass function** (j.p.m.f) of (X, Y).

Marginal probability mass functions: Let (X, Y) be a bivariate discrete random variable with joint probability mass function given by (x, y) . The marginal probability mass function of X and Y are given by

$$p_1(x_i) = \sum_{j=1}^{n} p\left(x_i, y_j\right) \text{ for } i = 1,2,3, \dots, m \text{ and}$$

$$p_2(y_j) = \sum_{i=1}^{m} p\left(x_i, y_j\right) \text{ for } i = 1,2,3, \dots, n$$

respectively.

Note: X and Y are independent if and only if $p(x_i, y_j) = p_1(x_i)p_2(x_j)\ \forall\ (i, j)$

Conditional probability mass functions: Let (X, Y) be a bivariate discrete random variable with joint probability mass function given by $p(x, y)$. The conditional probability mass function of X given $y = y_j$ and the conditional probability mass function of Y given $x = x_i$ are given by

$$p_{1|2}(x_i|\ y_j) = \frac{p(x_i, y_j)}{p_2(y_j)} \text{ for } i = 1,2,3, \dots, m \text{ and}$$
$$p_{2|1}(\ y_j|\ x_i) = \frac{p(x_i, y_j)}{p_1(x_i)}\ \ j = 1,2,3, \dots, n$$

respectively.

Example 1: A fair coin is tossed three times. Let X be a random variable that takes the value 0 if the first toss is a tail and the value 1 if the first toss is a head and Y be a random variable that defines the total number of heads in the three tosses. Then

i. **Determine the joint, marginal and conditional mass functions of X and Y.**
ii. **Are X and Y independent?**

Solution:

i. The sample space and values of X and Y are given in the following table:

Out comes in sample space	Value of X	Value of Y
HHH	1	3
HHT	1	2
HTH	1	2
HTT	1	1
THH	0	2
THT	0	1
TTH	0	1
TTT	0	0

Here X takes the values 0 and 1 and Y takes the values 0 ,1,2 and 3. Then the j.p.m.f of (x, y) is computed as below:

$$p(0,0) = P(X = 0, Y = 0) = P(\{TTT\}) = \frac{1}{8}$$

$$p(0,1) = P(X = 0, Y = 1) = P(\{THT, TTH\}) = \frac{2}{8} = \frac{1}{4}$$

$$p(0,2) = P(X = 0, Y = 2) = P(\{THH\}) = \frac{1}{8}$$

$$p(0,3) = P(X = 0, Y = 3) = 0$$

$$p(1,0) = P(X = 1, Y = 0) = 0$$

$$p(1,1) = P(X = 1, Y = 1) = P(\{HTT\}) = \frac{1}{8}$$

$$p(1,2) = P(X = 1, Y = 2) = P(\{ HTH, HHT\}) = \frac{2}{8} = \frac{1}{4}$$

$$p(1,3) = P(X = 1, Y = 3) = P(\{HHH\}) = \frac{1}{8}$$

The m.p.m.f of X is given by

$$p_1(0) = p(0,0) + p(0,1) + p(0,2) + p(0,3) = \frac{1}{8} + \frac{2}{8} + \frac{1}{8} + 0 = \frac{1}{2} \text{ and}$$

$$p_1(1) = p(1,0) + \; p(1,1) + p(1,2) + p(1,3) = 0 + \frac{1}{8} + \frac{2}{8} + \frac{1}{8} = \frac{1}{2}$$

The m.p.m.f of Y is given by

$$p_2(0) = p(0,0) + p(1,0) = \frac{1}{8} + 0 = \frac{1}{8}$$

$$p_2(1) = p(0,1) + p(1,1) = \frac{2}{8} + \frac{1}{8} = \frac{3}{8}$$

$$p_2(2) = p(0,2) + p(1,2) = \frac{1}{8} + \frac{2}{8} = \frac{3}{8}$$

$$p_2(3) = p(0,3) + p(1,3) = 0 + \frac{1}{8} = \frac{1}{8}$$

The conditional p.m.f of X given Y is computed below:

$p_{1|2}\ (0|0) = \frac{p(0,0)}{p_2(0)} = \frac{\frac{1}{8}}{\frac{1}{8}} = 1$, $p_{1|2}\ (1|0) = \frac{p(1,0)}{p_2(0)} = \frac{0}{\frac{1}{8}} = 0$

$p_{1|2}\ (0|1) = \frac{p(0,1)}{p_2(1)} = \frac{\frac{2}{8}}{\frac{3}{8}} = \frac{2}{3}$, $p_{1|2}\ (1|1) = \frac{p(1,1)}{p_2(1)} = \frac{\frac{1}{8}}{\frac{3}{8}} = \frac{1}{3}$

$p_{1|2}\ (0|2) = \frac{p(0,2)}{p_2(2)} = \frac{\frac{1}{8}}{\frac{3}{8}} = \frac{1}{3}$, $p_{1|2}\ (1|2) = \frac{p(1,2)}{p_2(2)} = \frac{\frac{2}{8}}{\frac{3}{8}} = \frac{2}{3}$

$p_{1|2}\ (0|3) = \frac{p(0,3)}{p_2(3)} = \frac{0}{\frac{1}{8}} = 0$, $p_{1|2}\ (1|3) = \frac{p(1,3)}{p_2(3)} = \frac{\frac{1}{8}}{\frac{1}{8}} = 1$

The conditional p.m.f. of Y given X is computed as below:

$p_{2|1}\ (0|0) = \frac{P(0,0)}{P_1(0)} = \frac{\frac{1}{8}}{\frac{1}{2}} = \frac{1}{4}$, $P_{2|1}\ (1|0) = \frac{p(0,1)}{p_1(0)} = \frac{\frac{1}{4}}{\frac{1}{2}} = \frac{1}{2}$

$p_{2|1}\ (2|0) = \frac{p(0,2)}{p_1(0)} = \frac{\frac{1}{8}}{\frac{1}{2}} = \frac{1}{4}$, $p_{2|1}\ (3|0) = \frac{p(0,3)}{p_1(0)} = \frac{0}{\frac{1}{2}} = 0$

$p_{2|1}\ (0|1) = \frac{p(1,0)}{p_1(1)} = \frac{0}{\frac{1}{2}} = 0$, $p_{2|1}\ (1|1) = \frac{p(1,1)}{p_1(1)} = \frac{\frac{1}{8}}{\frac{1}{2}} = \frac{1}{4}$

$p_{2|1}\,(2|1) = \frac{p(1,2)}{p_1(1)} = \frac{\frac{1}{4}}{\frac{1}{2}} = \frac{1}{2}$, $\qquad p_{2|1}\,(3|1) = \frac{p(1,3)}{p_1(1)} = \frac{\frac{1}{8}}{\frac{1}{2}} = \frac{1}{4}$

ii. Here $p(0,0) = \frac{1}{8}$, $p_1(0) = \frac{1}{2}$ and $p_2(0) = \frac{1}{8}$

Since $p(0,0) \neq p_1(0)p_2(0)$, X and Y are not independent .

Example 2 : The j.p.m.f. of (X, Y) is given by

$$p(x,y)=\begin{cases} k(2x+y) & \textit{for } x=1,2;\ y=1,2 \\ 0 & \text{otherwise} \end{cases}$$

where k is a constant

a. **Find the value of k.**
b. **Find marginal and conditional p.m.fs.**
c. **Are X and Y independent.**

Solution:

a. Since $p(x,y)$ is a j.p.m.f, $\sum_{x=1}^{2}\sum_{y=1}^{2} p(x,y)=1$

$$\sum_{x=1}^{2}\sum_{y=1}^{2} p(x,y)=k\sum_{x=1}^{2}\sum_{y=1}^{2}(2x+y)=k(3+4+5+6)=18k=1.$$

Thus $k = \frac{1}{18}$

b. The m.p.m.f. of X is given by

$$p_1(x)=\sum_{y=1}^{2} p(x,y)=\frac{1}{18}\sum_{y=1}^{2} 2x+y=\frac{1}{18}\left[(2x+1)+(2x+2)\right]=\frac{4x+3}{18}$$

Thus, $p_1(x) = \frac{4x+3}{18}$ for $x = 1,2$.

The m.p.m.f of Y is given by

$$p_2(y)=\sum_{x=1}^{2} p(x,y)=\frac{1}{18}\sum_{x=1}^{2} 2x+y=\frac{1}{18}\left[(2+y)+(4+y)\right]=\frac{2y+6}{18}=\frac{y+3}{9}$$

Thus, $p_2(y) = \frac{y+3}{9}$ for $y = 1,2$

The c.p.m.f. of X given Y is given by

$$p_{1|2}(x|y) = \frac{p(x,y)}{p_2(y)} = \frac{\frac{1}{18}(2x+y)}{\frac{1}{18}(2y+6)} = \frac{2x+y}{2y+6}$$

Thus, $p_{1|2}(x|y) = \frac{2x+y}{2y+6}$ for $x = 1,2$

The c.p.m.f. of Y given X is given by

$$p_{2|1}(y|x) = \frac{p(x,y)}{p_1(x)} = \frac{\frac{1}{18}(2x+y)}{\frac{1}{18}(4x+3)} = \frac{2x+y}{4x+3}$$

Thus, $p_{2|1}(y|x) = \frac{2x+y}{4x+3}$ for $y = 1,2$

c. Note that $p_1(x).p_2(y) = \frac{4x+3}{18} . \frac{y+3}{9} \neq p(x,y)$.
   Thus, X and Y are not independent.

Joint probability density function: Let (X,Y) be a bivariate continuous random variable. Let

$$P(a \le x \le b, c \le y \le d) = \int_a^b \int_c^d f(x,y)\,dx\,dy$$

for some real numbers a, b, c, d such that $a < b$ and $c < d$. Then

i. $f(x, y) \geq 0 \; \forall (x, y)$ and

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

and the function $f(x, y)$ is known as the **joint probability density function** of the bivariate continuous random variable (X, Y).

Marginal probability density function: Let (X, Y) be a bivariate continuous random variable with j.p.d.f. $f(x, y)$. The marginal probability density functions of X and Y are given by

$$f_1(x) = \int_{-\infty}^{\infty} f(x,y) dy \quad \text{and} \quad f_2(y) = \int_{-\infty}^{\infty} f(x,y) dx$$

respectively.

Note: X and Y are independent if and only if $f(x, y) = f_1(x). f_2(y)$

Conditional probability density functions: Let (X, Y) be a bivariate continuous random variable with j.p.d.f. $f(x, y)$. Let $f_1(x)$ and $f_2(y)$ be the m.p.d.fs of X and Y respectively.The conditional probability density function of X given Y and the conditional probability density function of Y given $X$are given by

$$f_{1|2}\,(x|y) = \frac{f(x,y)}{f_2(y)} \text{ and}$$

$$f_{2|1}\,(y|x) = \frac{f(x, y)}{f_1(x)}$$

respectively.

Cumulative distribution function: The cumulative distribution of a bivariate random variable (X, Y) is defined by

$$F(x, y) = P(X \leq x, Y \leq y) \text{ and}$$

$$F(x,y)=\begin{cases}\sum\limits_{t\le x}\sum\limits_{s\le y} p(t,s) & if\ (X,Y)\, is\, a\ \ d.r.v\ with\ j.p.m.f.\ p(x,y)\\ \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} f(t,s)dt\,ds & if\ (X,Y)\, is\, a\ \ c.r.v\ with\ j.p.d.f.\ f(x,y)\end{cases}$$

Properties of cumulative distribution function

1. $0 \le F(x,y) \le 1$
2. $F(\infty,\infty) = 1\ , F(-\infty,-\infty) = 0$
3. $P(a < X \le b, Y \le d) = F(b,d) - F(a,d)$ and
   $P(\ X \le b, c < Y \le d) = F(b,d) - F(b,c)$
4. $P(a < X \le b\ , c < Y \le d) = F(b,d) - F(a,d) - F(b,c) + F(a,c)$

Marginal cumulative distribution function: Let (X,Y) be a bivariate random variable with c.d.f. $F(x,y)$.The marginal cumulative distribution functions of X and Y are given by $F_1(\ x) = F(x,\infty)$ and $F_2(\ y) = F(\infty, y)$ respectively.

Note:

If (X,Y) is a bivariate continuous random variable with c.d.f. $F(x,y)$, then its j.p.d.f. is given by

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x\ \partial y}$$

Example 3: The j.p.d.f. of (X,Y) is given by

$$f(x,y)=\begin{cases}e^{-(x+y)} & for\ 0<x<\infty, 0<y<\infty\\ 0 & otherwise\end{cases}$$

a. **Find marginal p.d.fs of X and Y.**
b. **Are X and Y independent?**

Solution:

a. The m.p.d.f of X is given by

$$f_1(x)=\int\limits_0^{\infty} f(x,y)dy=\int\limits_0^{\infty} e^{-(x+y)}dy=e^{-x}\int\limits_0^{\infty} e^{-y}dy=e^{-x}.1=e^{-x}$$

$$= \begin{cases} e^{-x} \; for \; x \geq 0 \\ 0 \quad otherwise \end{cases}$$

The m.p.d.f. of Y is given by

$$f_2(y) = \int_0^{\infty} f(x,y)dx = \int_0^{\infty} e^{-(x+y)}dx = e^{-y}\int_0^{\infty} e^{-x}dx = e^{-y}.1 = e^{-y}$$

$$= \begin{cases} e^{-y} \; for \; y \geq 0 \\ 0 \quad otherwise \end{cases}$$

b. Since $f_1(x).f_2(y) = e^{-x}.e^{-y} = e^{-(x+y)} = f(x,y)$, X and Y are independent.

Example 4: The j.p.d.f. of (X,Y) is given by

$$f(x,y) = \begin{cases} xe^{-x(y+1)} & 0 < x < \infty, 0 < y < \infty \\ 0 & \text{otherwise} \end{cases}$$

a. **Determine the marginal and conditional p.d.fs**
b. **Are X and Y independent.**

Solution: The m.p.d.f of X is given by

$$f_1(x) = \int_0^{\infty} f(x,y)dy = \int_0^{\infty} xe^{-x(y+1)}dy = xe^{-x}\int_0^{\infty} e^{-xy}dy = xe^{-x}\left[\frac{e^{-xy}}{-x}\right]_{y=0}^{y=\infty} = e^{-x}$$

$$\Rightarrow f_1(x) = e^{-x} \;\; for \; 0 < x < \infty$$

The m.p.d.f. of Y is given by

$$f_2(y) = \int_0^{\infty} f(x,y)dx = \int_0^{\infty} xe^{-x(y+1)}dx = \left[\frac{x.e^{-x(y+1)}}{y+1}\right]_0^{\infty} + \frac{1}{y+1}\int_0^{\infty} e^{-x(y+1)}dx$$

(using integration by parts)

$$=0-\frac{1}{(y+1)^2}\left[e^{-x(y+1)}\right]_0^{\infty}=\frac{1}{(y+1)^2}$$

$$\Rightarrow f_2(y)=\frac{1}{(y+1)^2} \quad for\ 0<y<\infty$$

The conditional p.d.f of X given Y is given by

$$f_{1|2}(x|y)=\frac{f(x,y)}{f_2(y)}=\frac{x.e^{-x.(y+1)}}{\frac{1}{(y+1)^2}}=x(y+1)^2e^{-x.(y+1)}$$

$\Longrightarrow f_{1|2}(x|y)=x(y+1)^2e^{-x.(y+1)}$ for $0<x<\infty$

The conditional p.d.f of Y given X is given by

$$f_{2|1}(x|y)=\frac{f(x,y)}{f_1(x)}=\frac{x.e^{-x.(y+1)}}{e^{-x}}=x.e^{-xy}$$

$\Longrightarrow f_{2|1}(x|y)=x.e^{-xy}$for $0<x<\infty$

Note that $f_1(x).f_2(y)=e^{-x}.\frac{1}{(y+1)^2}\neq f(x,y)$. Hence, X and Y are not independent .

Example 5: The j.p.d.f of (X, Y) is given by $f(x, y) = kx^3y$ for $0 < x < 2,\ 0 < y < 1$.

a. **Find k**
b. **Find the m.p.d.fs of X and Y**
c. **Are X and Y independent.**

Solution:

a. We have $\int_0^2\int_0^1 f(x,y)dx\,dy = \int_0^2\int_0^1 kx^3y\,dx\,dy = k\int_0^2 x^3\left(\int_0^1 y\,dy\right)dx = k\int_0^2 x^3\frac{1}{2}dx$

$$= \frac{k}{2}\left[\frac{x^4}{4}\right]_0^2 = \frac{k}{8}\times 16 = 2k$$

Now , $\int_0^2\int_0^1 f(x,y)dx\,dy = 1 \Rightarrow 2k = 1 \Rightarrow k = \frac{1}{2}$

The j.p.d.f of (X, Y) is given by

$$f(x, y) = \frac{1}{2}x^3y \text{ for } 0 < x < 2,\ 0 < y < 1$$

The m.p.d.f of X is given by

$$f_1(x) = \int_0^1 f(x,y)dy = \frac{1}{2}x^3\int_0^1 y\,dy = \frac{1}{2}x^3\left[\frac{y^2}{2}\right]_0^1 = \frac{1}{4}x^3$$

$\Rightarrow f_1(x) = \frac{x^3}{4}$ for $0 < x < 2$.

The m.p.d.f. of Y is given by

$$f_2(y) = \int_0^2 f(x,y)dx = \frac{1}{2}y\int_0^2 x^3\,dx = \frac{1}{2}y\left[\frac{x^4}{4}\right]_0^2 = 2y$$

$\Rightarrow f_2(y) = 2y$ for $0 < y < 1$.

b. Note that $f_1(x)f_2(y) = \frac{x^3}{4} . 2y = \frac{x^3 y}{2} = f(x, y), \forall (x, y)$

Since $f_1(x)f_2(y) = f(x, y), \forall (x, y)$ X and Y are independent.

2.3

Mathematical Expectation

The term expectation is used for the process of averaging when a random variable is involved. It is the number used to locate the centre of the probability distribution (p.m.f or p.d.f) of a random variable. A probability distribution is described by certain satisfied measures which are computed using mathematical expectation (or expectation)

Let X be a random variable defined on a sample space S. Let $g(.)$ be a function of X such that $g(X)$ is a random variable. Then the **expected value of $g(X)$** is defined by

$$E\left(g(X)\right)=\begin{cases}\sum\limits_{x} g(x)\,p(x) \;\; if\ X\ is\, a\, d.r.v\, with\, p.m.f.\ p(x)\\ \int\limits_{-\infty}^{\infty} g(x)\,f(x)dx \;\; if\ X\ is\, a\, c.r.v\, with\, p.d.f.\ f(x)\end{cases} \quad \text{--------}(1)$$

provided these values exist.

Mean and moments:

i. Let $g(X) = X$. Then, by formula (1), **expected value of X** is defined by

$$E(X)=\mu=\begin{cases}\sum\limits_{x} x\,p(x) \;\; if\ X\ is\, a\, d.r.v\, with\, p.m.f.\ p(x)\\ \int\limits_{-\infty}^{\infty} x\,f(x)dx \;\; if\ X\ is\, a\, c.r.v\, with\, p.d.f.\ f(x)\end{cases}$$

Then $E(X)$ is called the **mean of the random variable X** and it is denoted by μ.

ii. Let $g(X) = (X - A)^r$ where A is an arbitrary constant and r is a non negative integer. Then the formula (1) gives

$$E(X-A)^r=\mu_r'=\begin{cases}\sum\limits_{x} (x-A)^r\,p(x) \;\; if\ X\ is\, a\, d.r.v\, with\, p.m.f.\ p(x)\\ \int\limits_{-\infty}^{\infty} (x-A)^r\,f(x)dx \;\; if\ X\ is\, a\, c.r.v\, with\, p.d.f.\ f(x)\end{cases}$$

The quantity $E(X-A)^r$ is called the $\boldsymbol{r^{th}}$ **moment about $\boldsymbol{A}$** and it is denoted by μ_r'. If $A=0$, then μ_r' are known as **Raw Moments**.

iii. Let $g(X)=\big(X-E(X)\big)^r=(X-\mu)^r$.Then the formula (1) gives

$$E(X-\mu)^r=\mu_r=\begin{cases}\sum\limits_{x}(x-\mu)^r\,p(x) & \text{if } X \text{ is a d.r.v with p.m.f. } p(x)\\ \int\limits_{-\infty}^{\infty}(x-\mu)^r\,f(x)dx & \text{if } X \text{ is a c.r.v with p.d.f. } f(x)\end{cases}$$

The function $E(X-\mu)^r$ is called the $\boldsymbol{r^{th}}$ **central moment of $\boldsymbol{X}$** and it is denoted by $\boldsymbol{\mu_r}$

iv. If $r=2$, then $\mu_2=\sigma^2=E(X-\mu)^2$ and it is known as the **variance of the random variable $\boldsymbol{X}$** and it is denoted by $\boldsymbol{V(X)}$ or $\boldsymbol{\sigma^2}$.

v. Mean (μ) and variance (σ^2) are important statistical measures of a probability distribution.

Example 1: Let $\boldsymbol{X}$ be a d.r.v with the p.m.f. given below:

$\boldsymbol{x}$	$\boldsymbol{-3}$	$\boldsymbol{6}$	$\boldsymbol{9}$
$\boldsymbol{p(x)}$	$\boldsymbol{\frac{1}{6}}$	$\boldsymbol{\frac{1}{2}}$	$\boldsymbol{\frac{1}{3}}$

Find $\boldsymbol{E(X)}$and $\boldsymbol{E(X^2)}$.

Solution:

$$E(X)=\sum_x x\,p(x)-3\times\frac{1}{6}+\;6\times\frac{1}{2}+9\times\frac{1}{3}=-\frac{1}{2}+3+3=\frac{11}{2}$$

$$E(X^2)=\sum_x x^2p(x)=9\times\frac{1}{6}+36\times\frac{1}{2}+81\times\frac{1}{3}=\frac{93}{2}$$

Example 2: Find the expectation of the number on a die when thrown.

Solution: Let X be the random variable representing the number on a die when thrown. Then X can take any one of the values 1,2,3,4,5,6 each with equal probability $\frac{1}{6}$. Hence

$$E(X) = \sum_x x\, p(x) = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6}$$

$$= \frac{1}{6}(1+2+3+4+5+6) = \frac{1}{6}.\frac{6.7}{2} = \frac{7}{2}$$

$$\Rightarrow E(X) = \frac{7}{2}$$

Example 3: Two unbiased dice are thrown. Find the expected values of the sum of numbers of points on them.

Solution: Define X is the sum of the numbers obtained on the two dice and $X = 2,3,4, \ldots, 12$ and its probability distribution is given by

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

$$E(X) = \sum x\, p(x) = 2 \times \frac{1}{36} + 3 \times \frac{2}{36} + 4 \times \frac{3}{36} + 5 \times \frac{4}{36} + 6 \times \frac{5}{36} +$$

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

$$= \frac{1}{36}(2+6+12+20+30+42+40+36+30+22+12) = \frac{252}{36} = 7$$

$$\Rightarrow E(X) = 7$$

Example 4: In four tosses of a coin, let X be the number of heads. Find the mean and variance of X.

Solution: The sample space S consists of $2^4 = 16$ outcomes and the following table gives the outcomes and the value of X for each outcome is

S.No	**Out come**	X
1	$TTTT$	0
2	$TTTH$	1
3	$TTHT$	1
4	$TTHH$	2
5	$THTT$	1
6	$THTH$	2
7	$THHT$	2
8	$THHH$	3
9	$HTTT$	1
10	$HTTH$	2
11	$HTHT$	2
12	$HTHH$	3
13	$HHTT$	2
14	$HHTH$	3
15	$HHHT$	3
16	$HHHH$	4

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

The p.m.f of X is given in the following table:

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

$$\mathrm{E(X)} = \sum_x x\, p(x) = 0 \times \frac{1}{16} + 1 \times \frac{4}{16} + 2 \times \frac{6}{16} + 3 \times \frac{4}{16} + 4 \times \frac{1}{6}$$

$$= \frac{1}{16}(0 + 4 + 12 + 12 + 4) = \frac{32}{16} = 2$$

$$\Longrightarrow \mathrm{E(X)} = 2$$

$$V(X) = E(X-2)^2 = \sum (x-2)^2\, p(x)$$

$$= (0-2)^2 \times \frac{1}{16} + (1-2)^2 \times \frac{4}{16} + (2-2)^2 \times \frac{6}{16} + (3-2)^2 \times \frac{4}{16} + (4-2)^2 \times \frac{1}{16}$$

$$= 4 \times \frac{1}{16} + 1 \times \frac{4}{16} + 0 \times \frac{6}{16} + 1 \times \frac{4}{16} + 4 \times \frac{1}{16} = \frac{1}{16}(4+4+4+4) = \frac{16}{16} = 1$$

$\Longrightarrow \mathrm{V(X)} = 1$

Example 5: Find the mean and variance of the random variable X, whose p.d.f is given by

$$f(x) = \begin{cases} \frac{1}{2}, & 0 < x < 2 \\ 0, & otherwise \end{cases}$$

Solution:

$$E(X) = \int_0^2 x.f(x)dx = \frac{1}{2}\int_0^2 x.dx = \frac{1}{2}\left[\frac{x^2}{2}\right]_0^2 = 1 - 0 = 1$$

$\Longrightarrow$ Mean of the random variable X is 1.

Variance $= E(X-1)^2 =$

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

$\Longrightarrow$ Variance of the random variable X is $\frac{1}{3}$

Example 6: Find the mean of the random variable X whose p.d.f. is given by

$$f(x) = \begin{cases} e^{-x}, & x \geq 0 \\ 0 \quad, & otherwise \end{cases}$$

Solution:

$$E(X) = \int_0^\infty x\, f(x) dx = \frac{1}{2}\int_0^2 x e^{-x} dx = \left[-xe^{-x}\right]_0^\infty + \int_0^\infty e^{-x} dx = 0 + \left[-e^{-x}\right]_0^\infty = 0+1=1$$

$\Longrightarrow E(X) = 1$

Theorems on Mathematical Expectation:

The following theorems are proved by assuming that the random variables are continuous. If the random variables are discrete, the proof remains the same except replacing integration by summation.

Theorem 1: If X is a random variable and a and b are constants then

$$\boldsymbol{E(aX + b) = aE(X) + b}.$$

Proof: Let X be a c.r.v with p.d.f. $f(x)$. Then

$E(aX + b) =$

$$\int_{-\infty}^{\infty} (ax+b)\, f(x).dx = a\int_{-\infty}^{\infty} x\, f(x)\, dx + b\int_{-\infty}^{\infty} f(x).dx = a\,E(X) + b \qquad \left(\because \int_{-\infty}^{\infty} f(x).dx = 1\right)$$

Corollary 1: If $b = 0$, then $E(aX) = aE(X)$

Corollary 2: If $X = 1$ and $b = 0$, then $E(a) = a$

Theorem 2: Addition Theorem of mathematical expectation.

If X and Y are random variables, then $E(X + Y) = E(X) + E(Y)$ provided all the expectations exist.

Proof: Let X and Y be continuous random variables with j.p.d.f. $f(x, y)$ and m.p.d.fs be $f_1(x)$ and $f_2(y)$ respectively. Then by definition,

$E(X) = \int_{-\infty}^{\infty} x\, f_1(x).dx$ and $E(Y) = \int_{-\infty}^{\infty} y.f_2(y).dy$

Now, $E(X + Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x+y)\, f(x,y).dx.dy$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x\, f(x,y)\, dxdy + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y\, f(x,y)\, dxdy$$

$$= \int_{-\infty}^{\infty} x \left[\int_{-\infty}^{\infty} f(x,y)\, dy \right] dx + \int_{-\infty}^{\infty} y \left[\int_{-\infty}^{\infty} f(x,y)\, dx \right] dy$$

$$= \int_{-\infty}^{\infty} x\, f_1(x)\, dx + \int_{-\infty}^{\infty} y\, f_2(y)\, dy = E(X) + E(Y)$$

$$\therefore E(X+Y) = E(X) + E(y)$$

Generalization: If $X_1, X_2, X_3 \ldots X_n$ are random variables, then $E(X_1 + X_2 + X_3 + \cdots + X_n) = E(X_1) + E(X_2) + E(X_3) \ldots + E(X_n)$ provided all the expectations exist.

Theorem 3: Multiplication Theorem of mathematical Expectations

If X and Y are independent random variables, then $E(X\,Y) = E(X)\,E(Y)$.

Proof: Let X and Y be continuous random variables with j.p.d.f. $f(x,y)$ and m.p.d.fs be $f_1(x)$ and $f_2(y)$ respectively. Then by definition,

$$E(X) = \int_{-\infty}^{\infty} x\, f_1(x)\, dx \text{ and } E(Y) = \int_{-\infty}^{\infty} y\, f_2(y)\, dy$$

Now, $E(X.Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (xy)\, f(x,y)\, dx\, dy$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x,y)\, f_1(x)\, f_2(y)\, dx\, dy \qquad (\because X \text{ and } Y \text{ are independent})$$

$$= \left(\int_{-\infty}^{\infty} x\, f_1(x)\, dx \right) \left(\int_{-\infty}^{\infty} y\, f_2(y)\, dy \right) = E(X)\, E(Y)$$

Generalization: If $X_1, X_2, X_3 \ldots X_n$ are independent random variables, then $E(X_1 X_2 X_3 \ldots . X_n) = E(X_1) E(X_2) E(X_3) \ldots E(X_n)$.

Theorem 4: Mathematical expectation of a linear combination of random variables.

Let $\boldsymbol{X_1, X_2, X_3 \dots X_n}$ be any $\boldsymbol{n}$ random variables and $\boldsymbol{a_1, a_2, a_3, \dots . a_n}$ be any $\boldsymbol{n}$ constants. Then

$$\boldsymbol{E(a_1X_1 + a_2X_2 + a_3X_3 + \cdots + a_nX_n) = a_1E(X_1) + a_2E(X_2) + a_3E(X_3) + \cdots + a_nE(X_n)}$$

provided all the expectations exist.

The proof follows using Theorem 1 and generalization of Theorem 2.

Theorem 5: $\boldsymbol{V(X) = E(X^2) - \left(E(X)\right)^2}$

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

$$= E\left[X^2 - 2XE(X) + \left(E(X)\right)^2\right]$$

$$= E(X^2) - 2E\big(X\,E(X)\big) + E\big(E(X)\big)^2$$

$= \; E(X^2) - 2E(X)E(X) + \left(E(X)\right)^2 \;\; \because E(X)$ is a constant and $E\big(E(X)\big) = E(X)$

$\Rightarrow V(X) = E(X^2) - 2\left(E(X)\right)^2 + \left(E(X)\right)^2$

$\Rightarrow V(X) = E(X^2) - \left(E(X)\right)^2$

Note: The formula is simple to use instead of $\boldsymbol{E\big(X - E(X)\big)^2}$.

Theorem 6: If $\boldsymbol{X}$ is a random variable, and $\boldsymbol{a}$ and $\boldsymbol{b}$ are constants, then $\boldsymbol{V(ax + b) = a^2\, V(X)}$.

Proof: Let $Y = aX + b$. Then $E(Y) = E(aX + b) = aE(X) + b$ and

$Y - E(Y) = a(X - E(X))$

$\Rightarrow E\big(Y - E(Y)\big)^2 = a^2E\big(X - E(X)\big)^2$

$\Rightarrow V(Y) = a^2V(X) \Rightarrow V(aX + b) = a^2V(X)$

Corollary 1: If $a = 0$, then $V(b) = 0$ $i.e.$, variance of a constant is zero.

Corollary 2: If $b = 0$, then $V(aX) = a^2V(X)$

Covariance: If X and Y are two random variables, then the **covariance** between them is defined by

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

$= E[XY - XE(Y) - YE(X) + E(X)E(Y)]$

$= E(XY) - E\left(\left(XE(Y)\right) - E(YE(X)\right) + E(E(X)E(Y))$

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

$= E(XY) - E(X)E(Y)$

Note:

1. If X and Y are independent , then $\text{Cov}(X, Y) = 0$
2. $\text{Cov}(aX, bY) = ab\text{Cov}(X, Y)$ where a and b are constants.
3. $\text{Cov}(X + a, Y + b) = \text{Cov}(X, Y)$.

Theorem 7: Variance of a linear combination of random variables.

Let $X_1, X_2, X_3, \dots, X_n$ be any n random variables and $a_1, a_2, a_3, \dots, a_n$ are n constants, then

$$V\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 V(X_i) + 2\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j Cov\left(X_i, X_j\right)$$

Proof:

Let $U = \sum_{i=1}^{n} a_i X_i$, then $E(U) = \sum_{i=1}^{n} a_i E\left(X_i\right)$ and $U - E(U) = \sum_{i=1}^{n} a_i\left(X - E\left(X_i\right)\right)$

$$\Rightarrow \left(U - E(U)\right)^2 =$$

$$\sum_{i=1}^{n} a_i^2\left(X_i - E\left(X_i\right)\right)^2 + 2\sum_{\substack{i=1 \\ i<j}}^{n}\sum_{j=1}^{n} a_i a_j\left(X_i - E\left(X_i\right)\right)\left(X_j - E\left(X_j\right)\right)$$

$$\Rightarrow E\left(U - E(U)\right)^2 =$$

$$\sum_{i=1}^{n} a_i^2 E\left(X_i - E\left(X_i\right)\right)^2 + 2\sum_{\substack{i=1 \\ i<j}}^{n}\sum_{j=1}^{n} a_i a_j E\left[\left(X_i - E\left(X_i\right)\right)\left(X_j - E\left(X_j\right)\right)\right]$$

$$\Rightarrow V\left(\sum_{i=1}^{n} a_i X_i\right) = V(U) = \sum_{i=1}^{n} a_i^2 V\left(X_i\right) + 2\sum_{\substack{i=1 \\ i<j}}^{n}\sum_{j=1}^{n} a_i a_j \operatorname{cov}\left(X_i, X_j\right)$$

Note:

1. **If $X_1, X_2, X_3, \dots, X_n$ are independent , then** $V\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 V\left(X_i\right)$
2. **If $a_1 = a_2 = 1$ and $a_3 = \cdots . = a_n = 0$, then**
   $\boldsymbol{V(X_1 + X_2) = V(X_1) + V(X_2) + 2Cov(X_1, X_2)}$

3. **If $a_1 = 1, a_2 = -1$ and $a_3 = \cdots . = a_n = 0$, then**
   $\boldsymbol{V(X_1 - X_2) = V(X_1) + V(X_2) - 2Cov(X_1, X_2)}$

4. **If X_1 and X_2 are independent , then $V(X_1 \pm X_2) = V(X_1) + V(X_2)$**

Example 7: The j.p.d.f. of X and Y is given by

$$f(x,y) = \begin{cases} 2 - x - y & ,0 < x < 1, 0 < y < 1 \\ 0 & , \quad otherwise \end{cases}$$

Find

i. **m.p.d.fs of X and Y**
ii. **c.p.d.fs of X and Y**
iii. **$V(X)$ and $V(Y)$**

iv. **Covariance between X and Y**

Solutions:

i. $$f_1(x) = \int_0^1 f(x,y)\,dy = \int_0^1 (2-x-y)\,dy = \left[2y - xy - \frac{y^2}{2}\right]_0^1 = 2 - x - \frac{1}{2} = \frac{3}{2} - x$$

$$f_1(x) = \begin{cases} \frac{3}{2} - x & ,0 < x < 1 \\ 0 & ,otherwise \end{cases}$$

Similarly $$f_2(y) = \begin{cases} \frac{3}{2} - y & ,0 < y < 1 \\ 0 & ,otherwise \end{cases}$$

ii. $f_{1|2}(x|y) = \frac{f(x,y)}{f_2(y)} = \frac{2-x-y}{\frac{3}{2}-y}, 0 < x, y < 1$

and $f_{2|1}(y|x) = \frac{f(x,y)}{f_1(x)} = \frac{2-x-y}{\frac{3}{2}-x},\ 0 < x, y < 1$

iii. $E(X) = \int_0^1 x f_1(x)\,dx = \int_0^1 x\left(\frac{3}{2} - x\right)dx = \frac{5}{12}$ and

$E(X^2) = \int_0^1 x^2 f_1(x)\,dx = \int_0^1 x^2\left(\frac{3}{2} - x\right)dx = \frac{1}{4}$

Thus $V(X) = E(X^2) - \left(E(X)\right)^2 = \frac{1}{4} - \frac{25}{144} = \frac{11}{144}$

Similarly $V(Y) = \frac{11}{144}$

iv. $E(XY) = \int_0^1\int_0^1 xy f(x,y)\,dxdy = \int_0^1\int_0^1 xy(2-x-y)\,dxdy = \frac{1}{6}$ (verify!)

$$\therefore Cov(X,Y) = E(XY) - E(X)E(Y) = \frac{1}{6} - \frac{5}{12}.\frac{5}{12} = -\frac{1}{144}$$

2.4

Discrete Probability Distributions

Modules 2.1 and 2.2 deal with general properties of random variables**.** Random variables with special probability distributions are encountered in different fields of *science* and *engineering*. Some specific **discrete probability distributions** are discussed in this module and some specific **continuous probability distributions** are discussed in the next module 2.5.

Discrete Uniform Distribution: A r.v.X is said to have a **discrete uniform distribution** over the range $[1, n]$, if its p.m.f. is given by

$$p(x) = P(X = x) = \begin{cases} \frac{1}{n} & , \quad x = 1,2,\dots,n \\ 0 & , \quad \text{otherwise} \end{cases}$$

Notation: $X \sim U(n)$,read as X follows discrete uniform distribution with parameter n.

Note: If all possible values of a r.v. are equally likely, then this distribution is used.

Example 1: If an unbiased coin is tossed once and X is equal to number of heads, then $X = 0,1$ and

$$P(X = 0) = P(X = 1) = \frac{1}{2} \text{ and } X \sim U(2).$$

Example 2: If an unbiased die is thrown once and X is equal to number on the die, then $x = 1,2,3,4,5,6$ and $P(X = i) = \frac{1}{6}$ for $i = 1,2,3,4,5,6$ and $X \sim U(6)$.

Mean and Variance: We have $E(X) = \frac{1}{n}\sum_{i=1}^{n} i = \frac{n+1}{2}$

and $E\left(X^2\right) = \frac{1}{n}\sum_{i=1}^{n} i^2 = \frac{(n+1)(2n+1)}{6}$

Thus $V(X) = E\left(X^2\right) - \left(E(X)\right)^2 = \frac{(n+1)(n-1)}{12}$

Bernoulli Experiment: A random experiment whose outcomes are of two types, **success (S)** and **failure (F),** occurring with probabilities p and $q(= 1 - p)$ respectively, is called a **Bernoulli experiment**.

Conducting a Bernoulli experiment once is known as **Bernoulli trial**. Note that p and q are same in each trial and outcomes of different trials are independent.

Bernoulli distribution: In a Bernoulli experiment, if a r.v.X is defined such that it takes value 1 with probability p when S occurs and 0 with probability q when F occurs, then we say that X follows Bernoulli distribution and its p.m.f. is given by

$$p(x) = P(X = x) = \begin{cases} p^x q^{1-x} & , \quad x = 0,1 \\ 0 & , \quad \text{otherwise} \end{cases}$$

Examples:

1) Tossing of a coin (results a head or tail)
2) Performance of a student in an examination (results pass or failure)
3) Sex of an unborn child (results female or male)

Mean and Variance:

Mean $= \mu = E(X) = 0 \times q + 1 \times p = p$

and $E(X^2) = 0^2 \times q + 1^2 \times p = p$

$\therefore$ Variance $= \sigma^2 = E(X^2) - \left(E(X)\right)^2 = p - p^2 = p(1 - p) = pq$

Binomial Distribution: Suppose we conduct n independent Bernoulli trials and we define

$$X = \text{number of successes in } n \text{ trials.}$$

Then X is a discrete random variable and it takes the values 0,1,2,…,n.

Derivation of $P(X = x)$: Note that $X = x$ means that there are x successes and $(n - x)$ failures in n trials in a specified order (say) SSFSFFFS … FSF.

Since outcomes of different trials are independent, by Multiplication Theorem, we have

$$P(SSFSFFFS \ldots FSF) = P(S) \cdot P(S) \cdot P(F) \cdot P(S) \cdot P(F) \cdot P(F) \cdot P(F) \cdot P(S) \cdot \ldots P(F) \cdot P(S) \cdot P(F)$$

$$= p\ p\ q\ p\ q\ q\ q\ p \ldots q\ p\ q$$

$$= \underbrace{p \cdot p \cdot \ldots \cdot p}_{(x\ times)} \cdot \underbrace{q \cdot q \cdot \ldots \cdot q}_{(n-x\ times)} = p^x q^{n-x}$$

But x successes in n trials can occur in $\binom{n}{x}$ orders and the probability for each of these orders is same, viz., $p^x q^{n-x}$. Hence by addition theorem of probability

$$p(x) = P(X = x) = \binom{n}{x} p^x q^{n-x}$$

Definition: A r.v. X is said to follow a **binomial distribution** with parameters n and p if its p.m.f. is given by

$$p(x) = P(X = x) = \begin{cases} \binom{n}{x} p^x q^{n-x} & , \quad x = 0,1,2, \ldots, n, 0 < p < 1, q = 1 - p \\ 0 & , \quad \text{otherwise} \end{cases}$$

Notation: $X \sim B(n, p)$. Read as X follows binomial distribution with parameters n and p.

Real life examples:

1) Number of heads in n tosses of a coin
2) Number of boys in a family of n children
3) Number of times hitting a target in n attempts

Note:

1.

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

2. The c.d.f. of X is given by

$$F(x) = P(X \leq x) = \sum_{k=0}^{x} \binom{n}{k} p^k q^{n-k} \ , x = 0,1,2, \dots , n$$

Example 1: Four fair coins are tossed. If the outcomes are assumed to be independent, then find the p.m.f. and c.d.f. of the number of heads obtained.

Solution: Let X be the no. of heads in tossing 4 coins.

Then $X \sim B\left(4, \frac{1}{2}\right)$ where $p = P(head) = \frac{1}{2}$.

Thus $p(x) = P(X = x) = \binom{4}{x} \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{4-x}$

$$= \binom{4}{x} \left(\frac{1}{2}\right)^4 = \binom{4}{x} \left(\frac{1}{16}\right) \text{ for } x = 0,1,2,3,4.$$

Then $p(0) = \binom{4}{0} \left(\frac{1}{16}\right) = \frac{1}{16}$

$p(1) = \binom{4}{1} \left(\frac{1}{16}\right) = \frac{4}{16}$

$p(2) = \binom{4}{2} \left(\frac{1}{16}\right) = \frac{6}{16}$

$p(3) = \binom{4}{3} \left(\frac{1}{16}\right) = \frac{4}{16}$

$p(4) = \binom{4}{4} \left(\frac{1}{16}\right) = \frac{1}{16}$

The p.m.f $p(x)$ and c.d.f $F(x)$ are given in the following table.

x	0	1	2	3	4
$p(x)$	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$
$F(x)$	$\frac{1}{16}$	$\frac{5}{16}$	$\frac{11}{16}$	$\frac{15}{16}$	1

Example 2: A and B play a game in which their chances of winning are in the ratio $3:2$. Find A's chance of winning at least three games out of the five games played.

Solution:

Define $X =$ No. of games A winning out of 5.

Here $p = P(A\ winning) = \frac{3}{5}$ and $n = 5$ and $X \sim B\left(5, \frac{3}{5}\right)$. Thus,

$$p(x) = P(X = x) = \binom{5}{x}\left(\frac{3}{5}\right)^x\left(\frac{2}{5}\right)^{5-x} \text{ for } x = 0,1,\ldots,5.$$

Required to find:

$$P(A \text{ winning at least 3 out of 5 games}) = P(X \geq 3)$$

$$= P(X = 3) + P(X = 4) + P(X = 5)$$

$$= \binom{5}{3}\left(\frac{3}{5}\right)^3\left(\frac{2}{5}\right)^2 + \binom{5}{4}\left(\frac{3}{5}\right)^4\left(\frac{2}{5}\right)^1 + \binom{5}{5}\left(\frac{3}{5}\right)^5$$

$$= \frac{3^3}{5^5}[10 \times 4 + 5 \times 3 \times 2 + 1 \times 9]$$

$$= \frac{27 \times (40 + 30 + 9)}{3125} = 0.68$$

Example 3: The probability of a man hitting a target is $\frac{1}{4}$.

(i) **If he fires 7 times, what is the probability of his hitting the target at least twice?**

(ii) **How many times must he fire so that the probability of his hitting the target at least once is greater than $\frac{2}{3}$?**

Solution: Let X be the no. of times a man hitting the target in 7 fires. Here $p = P(man\ hitting\ the\ target) = \frac{1}{4}$ and $n = 7$. Then $X \sim B\left(7, \frac{1}{4}\right)$ and

$$p(x) = \binom{7}{x}\left(\frac{1}{4}\right)^x\left(\frac{3}{4}\right)^{7-x} \text{ for } x = 0,1,2,\ldots,7.$$

(i) $P(at\ least\ two\ hits) = P(X \geq 2) = 1 - [P(X = 0) + P(X = 1)]$

$$= 1 - [p(0) + p(1)]$$

$$= 1 - \left\{ \binom{7}{0} \left(\frac{1}{4}\right)^0 \left(\frac{3}{4}\right)^7 + \binom{7}{1} \left(\frac{1}{4}\right) \left(\frac{3}{4}\right)^6 \right\} = \frac{4547}{8192} = 0.55$$

(ii) Find n such that $P(X \geq 1) > \frac{2}{3}$

$$\Longrightarrow 1 - P(X = 0) > \frac{2}{3}$$

$$\Longrightarrow -1 + P(X = 0) < -\frac{2}{3}$$

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

$$\Longrightarrow \binom{n}{0} \left(\frac{1}{4}\right)^0 \left(\frac{3}{4}\right)^n < \frac{1}{3}$$

$$\Longrightarrow \left(\frac{3}{4}\right)^n < \frac{1}{3}$$

$$\Longrightarrow n[\log 3 - \log 4] < \log 1 - \log 3$$

$$\Longrightarrow n[\log 4 - \log 3] > \log 3$$

$\Longrightarrow n > \frac{\log 3}{\log 4 - \log 3} = 3.8$,since n cannot be fractional, the required number of shots is 4.

Mean of Binomial Distribution:

$$\mu = E(X) = \sum_{x=0}^{n} x\, p(x) = \sum_{x=0}^{n} x \binom{n}{x} p^x q^{n-x}$$

$$= \sum_{x=1}^{n} x \left(\frac{n}{x}\right) \binom{n-1}{x-1} p^x q^{n-x} = np \sum_{x=1}^{n} \binom{n-1}{x-1} p^{x-1} q^{n-x}$$

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

$$\Longrightarrow \mu = np$$

Variance of Binomial Distribution:

$\sigma^2 = V(X) = npq$ **(See $\boldsymbol{P_1}$ for proof)**

Example 4: One hundred balls are tossed into 50 boxes. What is the expected number of balls in the tenth box.

Solution: If we think of the balls tossed as Bernoulli trials in which a success is defined as getting a ball in the tenth box, then $p = \frac{1}{50}$. If X denotes the number of balls that go into the tenth box.

Then $X \sim B\left(100, \frac{1}{50}\right)$ and $E(X) = np = 100 \times \frac{1}{50} = 2$.

Example 5: The mean and variance of binomial distribution are $\mathbf{4}$ and $\frac{\mathbf{4}}{\mathbf{3}}$ respectively. Find $\boldsymbol{P(X \geq 1)}$.

Solution: Here $X \sim B(n, p)$. But $np = 4$ and $np(1-p) = \frac{4}{3}$.

Hence $4(1-p) = \frac{4}{3} \Longrightarrow 1 - p = \frac{1}{3} \Longrightarrow p = \frac{2}{3}$ and $n = \frac{4}{p} = 4 \times \frac{3}{2} = 6$.

Thus, $X \sim B\left(6, \frac{1}{3}\right)$ and hence $P(X \geq 1) = 1 - P(X = 0)$

$$= 1 - \binom{6}{0}\left(\frac{2}{3}\right)^0\left(\frac{1}{3}\right)^6 = 1 - \left(\frac{1}{3}\right)^6$$

Poisson Distribution:

If $n \to \infty$ and $p \to 0$ such that $\lambda = np$ fixed, then $\binom{n}{x} p^x (1-p)^{n-x} \longrightarrow \frac{e^{-\lambda}\lambda^x}{x!}$ which is the p.m.f. of Poisson distribution (**See $\boldsymbol{P_2}$**). Thus the p.m.f. of **Poisson distribution** is obtained as the limit of p.m.f. of **binomial distribution.**

Definition: A r.v.X is said to follow a **Poisson distribution** with parameter λ if its p.m.f. is given by

$$p(x) = P(X = x) = \begin{cases} \dfrac{e^{-\lambda}\lambda^x}{x!} & , \quad x = 0,1,2,\ldots; \lambda > 0 \\ 0 & , \quad \text{otherwise} \end{cases}$$

Notation: Read $X \sim P(\lambda)$ as: X follows poisson distribution with parameter λ.

Real life examples

1) Number of defectives in a packet of 100 blades.
2) Number of telephone calls received at a particular telephone exchange in some unit of time.
3) Number of print mistakes in a page of a book.
4) The number of fragments received by a surface area 'A' from a fragment atom bomb.
5) The emission of radio active (alpha) particles.
6) Number of air accidents in some unit of time.

Note :

1. $$\sum_{x=0}^{\infty} p(x) = \sum_{x=0}^{\infty} \frac{e^{-\lambda}\lambda^{x}}{x!} = e^{-\lambda}\sum_{x=0}^{\infty}\frac{\lambda^{x}}{x!} = e^{-\lambda}e^{\lambda} = 1\left(\because e^{x} = \sum_{i=0}^{\infty}\frac{x^{i}}{i!}\right)$$

2. The c.d.f. of X is given by

$$F(x) = P(X \leq x) = \sum_{k=0}^{x} \frac{e^{-\lambda}\lambda^{k}}{k!}$$

Example 6: Messages arrive at a switchboard in a Poisson manner at an average rate of six per hour. Find the probability for each of the following events:

a) **Exactly two messages arrive within one hour.**
b) **No message arrives within one hour.**
c) **At least three messages arrive within one hour.**

Solution: Let X be the r.v. that denotes the number of messages arriving at the switchboard within a one-hour interval. Then $X \sim P(6)$ and its p.m.f is given by

$$P(X = x) = p(x) = \frac{e^{-6}6^{x}}{x!} \text{ for } x = 0,1,2,\ldots$$

a) $P(X=2)=\frac{e^{-6}6^2}{2!}=\frac{36}{2}e^{-6}=18e^{-6}$.

b) $P(X=0)=\frac{e^{-6}6^0}{0!}=e^{-6}$.

c) $P(X\geq 3)=1-\{P(X=0)+P(X=1)+P(X=2)\}$

$$=1-\left\{\frac{e^{-6}6^0}{0!}+\frac{e^{-6}6^1}{1!}+\frac{e^{-6}6^2}{2!}\right\}=1-e^{-6}\{1+6+18\}=1-25e^{-6}$$

Example 7: In a book of 520 pages, 390 typo-graphical errors occur. Assuming Poisson law for the number of errors per page, find the probability that a random sample of 5 pages will contain no error.

Solution:

$\lambda=$ Average number of typo-graphical errors/page$=\frac{390}{520}=0.75$

Let $X=$ Number of errors per page

Then $X\sim P(0.75)$ and $p(x)=P(X=x)=\frac{e^{-0.75}(0.75)^x}{x!}$

P(No error)$=P(X=0)=p(0)=e^{-0.75}$

P(A random sample of 5 pages contain no error)$=[p(0)]^5=[e^{-0.75}]^5=e^{-3.75}$

Mean of Poisson Distribution : The mean of poisson distribution is given by

$$\mu=E(X)=\sum_{x=0}^{\infty}x\,p(x)=\sum_{x=1}^{\infty}\frac{x\,e^{-\lambda}\lambda^x}{x!}=\lambda e^{-\lambda}\sum_{x=1}^{\infty}\frac{\lambda^{x-1}}{(x-1)!}=\lambda e^{-\lambda}e^{\lambda}=\lambda$$

$\Rightarrow \mu=\lambda$

Variance of Poisson Distribution:

$$E(X^2)=\sum_{x=0}^{\infty}x^2P(x)=\sum_{x=0}^{\infty}[x(x-1)+x]p(x)$$

$$= \sum_{x=2}^{\infty} x(x-1)p(x) + \sum_{x=1}^{\infty} xp(x) = \sum_{x=2}^{\infty} x(x-1)\frac{e^{-\lambda}\lambda^x}{x!} + \lambda$$

$$= e^{-\lambda}\lambda^2 \sum_{x=2}^{\infty} \frac{\lambda^{x-2}}{(x-2)!} + \lambda = e^{-\lambda}\lambda^2 e^{\lambda} + \lambda$$

$\Longrightarrow E(X^2) = \lambda^2 + \lambda$

The variance of Poisson distribution is given by

$$V(X) = E(X^2) - \big(E(X)\big)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$$

$\Longrightarrow V(X) = \lambda$

Note that *for Poisson distribution, mean and variance are equal.*

Example 8: If X and Y are independent Poisson variates such that $P(X = 1) = P(X = 2)$ and $P(Y = 2) = P(Y = 3)$, find the variance of $X - 2Y$.

Solution: Let $X \sim P(\lambda)$ and $Y \sim P(\mu)$. Then

$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$ for $x = 0,1,2, \ldots;\ \lambda > 0$ and

$P(Y = y) = \frac{e^{-\mu}\mu^y}{y!}$ for $y = 0,1,2, \ldots;\ \mu > 0.$

Since $P(X = 1) = P(X = 2)$; $\lambda e^{-\lambda} = \frac{e^{-\lambda}\lambda^2}{2} \Longrightarrow \lambda^2 - 2\lambda = 0$

$$\Longrightarrow \lambda(\lambda - 2) = 0 \Longrightarrow \lambda = 0, 2 \Longrightarrow \lambda = 2\ (\because \lambda = 0 \text{ is not admissible})$$

Since $P(Y = 2) = P(Y = 3)$, then $\frac{e^{-\mu}\mu^2}{2} = \frac{e^{-\mu}\mu^3}{6} \Longrightarrow \mu^3 - 3\mu^2 = 0$

$$\Longrightarrow \mu^2(\mu - 3) = 0 \Longrightarrow \mu = 0,3 \Longrightarrow \mu = 3.$$

$V(X - 2Y) = 1^2\, V(X) + (-2)^2\, V(Y) = \lambda + 4\mu = 2 + 4 \times 3 = 2 + 12 = 14$

Negative Binomial (or Pascal) Distribution:

Let X denote the number of failures before the r^{th} success in a sequence of Bernoulli trials. Then the number of trials required is $X + r$.

Derivation of $P(X = x)$:

In $x + r$ trials, the last trial must be a success whose probability is p. In the remaining $(x + r - 1)$ trials, we must have $(r - 1)$ successes whose probability is $\binom{x+r-1}{r-1} p^{r-1} q^x$ (Using binomial distribution).

Thus, by multiplication theorem, we have

$$p(x) = P(X = x) = \binom{x+r-1}{r-1} p^{r-1} q^x \cdot p = \binom{x+r-1}{r-1} p^r q^x$$

Definition: A random variable X is said to follow a **Negative binomial distribution**(NBD) with parameters r and p if its p.m.f is given by

$$p(x) = P(X = x) = \begin{cases} \binom{x+r-1}{r-1} p^r q^x & , \quad x = 0,1,2,\ldots \\ 0 & , \quad \text{otherwise} \end{cases}$$

Notation: $X \sim NB(r, p)$.

Note:

1. $\binom{x+r-1}{r-1} = \binom{x+r-1}{x} \quad \left(\because \binom{n}{r} = \binom{n}{n-r}\right)$

$$= \frac{(x+r-1)(x+r-2)\ldots(r+1)r}{x!}$$

$$= (-1)^r \frac{(-r)(-r-1)\ldots(-r-x+2)(-r-x+1)}{x!} = (-1)^x \binom{-r}{x}$$

Thus, the p.m.f. of NBD can be written as

$$p(x) = \begin{cases} \binom{-r}{x} p^r (-q)^x & , \quad x = 0,1,2,\ldots \\ 0 & , \quad \text{otherwise} \end{cases}$$

Further , it is the $(x+1)^{th}$ term in the expansion of $p^r(1-q)^{-r}$, a binomial expansion with negative index. Therefore, the distribution is known as negative binomial distribution.

2. $\sum_{x=0}^{\infty} p(x) = p^r \sum_{x=0}^{\infty} \binom{-r}{x} (-q)^x = p^r (1-q)^{-r} = p^r p^{-r} = 1$

Mean of NBD: The mean of NBD is given by

$$\mu = E(x) = \sum_{x=0}^{\infty} x\, p(x) = \sum_{x=1}^{\infty} x\, p(x) = \sum_{x=1}^{\infty} x \binom{-r}{x} p^r (-q)^x$$

$$= p^r \sum_{x=1}^{\infty} x \left(\frac{-r}{x}\right) \binom{-r+1}{x-1} (-q)^x = (-r)(p^r)(-q) \sum_{x=1}^{\infty} \binom{-r+1}{x-1} (-q)^{x-1}$$

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

$$= rqp^r p^{-(r+1)} = \frac{rq}{p}$$

Variance of NBD: $\sigma^2 = \frac{rq}{p^2}$ **(See P_3 for proof)**

3. Notice that $\frac{\mu}{\sigma^2} = p < 1$ and this implies that mean is smaller than variance in NBD.
4. If $Y =$ Number of trials required to get r^{th} success, then $Y = X + r$ and

$$P(Y = y) = P(X + r = y) = P(X = y - r) = \binom{y-1}{r-1} p^r q^{y-r}$$

for $y = r, r+1, \ldots$ and $E(y) = E(X) + r = \frac{rq}{p} + 1$ and $V(Y) = V(X) = \frac{rq}{p^2}$

Real life examples

1) Number of tails before the third head.
2) Number of girls before the second son.
3) Number of non-defectives before the third defective.

Example 9: Find the probability that there are two daughters before the second son in a family when probability of a son is 0.5.

Solution: Let $X =$ the number of daughters before second son

Then $P(X = x) = \binom{x+2-1}{2-1}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^x = \binom{x+1}{1}\left(\frac{1}{2}\right)^{x+2}$

and $P(X = 2) = \binom{3}{1}\left(\frac{1}{2}\right)^4 = \frac{3}{16}$

Example 10: An item is produced in large numbers. The machine is known to produce 5% defectives. A quality control inspector is examining the items by taking them at random. What is the probability that at least 4 items are to be examined in order to get 2 defectives?

Solution: Let $Y =$ No. of items to be examined in order to get 2 defectives. Here $p =$(defective)$= \frac{5}{100} = 0.05$.

Then $P(Y = y) = \binom{y-1}{2-1}(0.05)^2(0.95)^{y-2}$

$\Rightarrow P(Y = y) = (y-1)(0.05)^2(0.95)^{y-2}$

We want to find

$$P(Y \geq 4) = 1 - \sum_{y=2}^{3} P(Y = y) = 1 - \{P(Y = 2) + P(Y = 3)\}$$

$$= 1 - \{(0.05)^2 + 2(0.05)^2(0.95)\} = 0.9928$$

Geometric distribution:

Let X denotes the number of failures before the first success in a sequence of Bernoulli trials. Then the required number of trials is $X + 1$.

Geometric distribution is a particular case of negative binomial distribution with $r = 1$.

Definition: A random variable X is said to follow a **Geometric distribution** (GD) with parameter p if its p.m.f. is given by

$$p(x) = P(X = x) = \begin{cases} p \cdot q^x & for \;\; x = 0,1,2,\dots \\ 0\,, & otherwise \end{cases}$$

Notation: $X \sim GD(r)$

Mean and variance of GD

$\mu = \frac{q}{p}$ and $\sigma^2 = \frac{q}{p^2}$ (take $r = 1$ in μ and σ^2 of NBD)

Note: Let $Y =$ Number of trials required to get first success, then $Y = X + 1$ and

$$P(Y = y) = P(X + 1 = y) = P(X = y - 1) = \begin{cases} p\, q^{y-1} & for \;\; y = 1,2,3,\dots \\ 0\,, & otherwise \end{cases}$$

Further, $E(Y) = E(X) + 1 = \frac{1}{p}$ and $V(Y) = V(X) = \frac{q}{p^2}$.

Real life examples

1) Number of tails before the third head
2) Number of girls before the second son
3) Number of non-defectives before the first defective

Example 11: find the probability that there are two daughters before the first son in a family where probability of a son is 0.5 .

Solution: Let $X =$ Number of daughters before the first son.

Then $P(X = x) = \left(\frac{1}{2}\right) \cdot \left(\frac{1}{2}\right)^x = \left(\frac{1}{2}\right)^{x+1}$ for $x = 0,1,2,\dots$

and $P(X = 2) = \left(\frac{1}{2}\right)^{2+1} = \left(\frac{1}{2}\right)^3 = \frac{1}{8}$

Hyper geometric Distribution:

Consider an urn with N balls, M of which are white and $N - M$ are red. Suppose we draw a sample of n balls at random with replacement. Let X denote the number of white balls in the sample. Then $X \sim B(n, p)$ where $p = \frac{M}{N}$ which remains same for all trials and outcomes of different trials are independent. The p.m.f of X is given by $P(X = x) = \binom{n}{x}\left(\frac{M}{N}\right)^x \left(1 - \frac{M}{N}\right)^{n-x}$ for $i = 1,2, \dots, n$.

If the sample is selected without replacement, p is not same for all trials and outcomes of different trials are not independent and hence binomial distribution cannot be applied.

Derivation of $P(X = x)$: The number of all possible samples without replacement $= \binom{N}{x}$.

The number of samples in which there are x white balls and $(n - x)$ red balls $= \binom{M}{x}\binom{N - M}{n - x}$.

Thus $p(x) = P(X = x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$.

Definition: A random variable X is said to follow the **hyper geometric distribution** if its p.m.f is given by

$$p(x) = P(X = x) = \begin{cases} \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} & , \ x = 0,1, \dots \min(n, M) \\ 0 & , \qquad otherwise \end{cases}$$

Example 12: A bag contains 4 white balls and 3 green balls. Three balls are drawn. What is the probability that 2 are white.

Solution: $N = 4 + 3 = 7, \ M = 4, n = 3$

$X =$ Number of white balls and

$$P(X = 2) = \frac{\binom{4}{2}\binom{3}{1}}{\binom{7}{3}} = \frac{4\times3\times3\times6}{2\times7\times6\times5} = \frac{18}{35}$$

Note: $$\sum_{x=0}^{\min(n,M)} p(x) = \begin{cases} \sum_{x=0}^{n} \dfrac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} = \dfrac{\binom{N}{n}}{\binom{N}{n}} = 1 \; if \; \min(n,M) = n \\ \sum_{x=0}^{n} \dfrac{\binom{M}{x}\binom{N-M}{M-x}}{\binom{N}{M}} = \dfrac{\binom{N}{M}}{\binom{N}{M}} = 1 \; if \; \min(n,M) = M \end{cases}$$

Mean and variance of Hyper geometric distribution:

The mean is given by $\mu = \frac{nM}{N}$ and variance is given by $\sigma^2 = \frac{NM(N-M)(N-n)}{N^2(N-1)}$ if $\min(n, M) = n$ **(For proof, see $\boldsymbol{P_4}$)**.

2.5

Continuous Probability Distributions

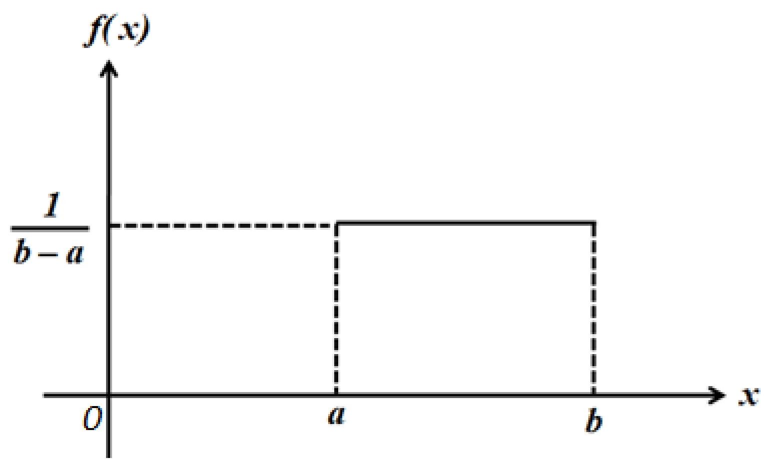
The **continuous probability distributions** are used in a number of applications in *engineering*. For example in *error analysis*, given a set of data or probability distribution, it is possible to estimate the probability that a measurement (temperature, pressure, flow rate) will fall within a desire range, and hence determine how reliable an instrument or piece of equipment is. Also, one can calibrate an instrument (ex. Temperature sensor) from the manufacturer on a regular basis and use a probability distribution to see of the variance in the instruments' measurements increases or decreases over time.

Uniform Distribution

A continuous random variable (c. r. v.) X is said to have a uniform distribution over the interval $[a, b]$ if its p. d. f. is given by

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & a \le x \le b \\ 0, & otherwise \end{cases}$$

Notation: $\boldsymbol{X \sim U}(a, b)$. Read as X *follows uniform distribution with parameters* a *and* b. It is used to model events that are equally likely to occur at any time within a given time interval. The plot of p. d. f. is given below:



The cumulative distribution function (c. d. f.) of X is given by

$$F(x) = p(X \leq x) = \begin{cases} 0 \quad , & x < a \\ \dfrac{x-a}{b-a} \quad , & a \leq x \leq b \\ 1 \quad , & x \geq b \end{cases}$$

The mean of X is given by

$$\mu = E(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx = \int_a^b x\, f(x)\, dx = \int_a^b \frac{x}{b-a}\, dx$$

$$= \left[\frac{x^2}{2(b-a)}\right]_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{(b-a)(b+a)}{2(b-a)} = \frac{b+a}{2} \implies \mu = \frac{b+a}{2}$$

Now, $E(X^2) = \int_{-\infty}^{\infty} x^2\, f(x)\, dx = \int_a^b x^2\, f(x)\, dx = \int_a^b \frac{x^2}{b-a}\, dx$

$$= \left[\frac{x^3}{3(b-a)}\right]_a^b = \frac{b^3-a^3}{3(b-a)} = \frac{(b-a)(b^2+ab+a^2)}{3(b-a)} = \frac{b^2+ab+a^2}{3}$$

$$\implies E(X^2) = \frac{b^2+ab+a^2}{3}$$

Thus, the variance of X is given by $\sigma^2 = E(X^2) - \left(E(X)\right)^2$

$$= \frac{b^2+ab+a^2}{3} - \frac{b^2+2ab+a^2}{4}$$

$$= \frac{b^2-2ab+a^2}{12}$$

$$\implies \sigma^2 = \frac{(b-a)^2}{12}$$

Example 1: The time that a professor takes to grade a paper is uniformly distributed between 5 *minutes* and 10 *minutes*. Find the mean and variance of the time the professor takes to grade a paper.

Solution: Let X denotes the time the professor takes to grade a paper. Then $X \sim U(5, 10)$.

$$\mu = E(X) = \frac{10+5}{2} = 7.5 \text{ and } \sigma^2 = V(X) = \frac{(10-5)^2}{12} = \frac{25}{12}\ (minutes)^2$$

Normal Distribution

The normal distribution was first discovered by **De – Movire** and **Laplace** as the limiting form of Binomial distribution. Through a historical error it was credited to Gauss who first made reference to it as the distribution of errors in Astromy. Gauss used the normal curve to describe theory of accidental errors of measurements involved in the calculation of orbits of heavenly bodies.

Definition: A c. r. v. X is said to have a **normal distribution** with parameters μ and σ^2 if its p. d. f. is given by

$$\boldsymbol{f(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}}, -\infty < x < \infty, -\infty < \mu < \infty, \sigma > 0$$

The c. d. f. of X is given by

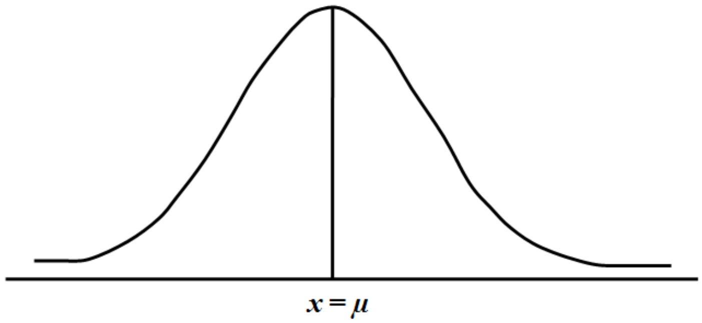
$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} exp\left\{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right\} dt$$

Notation: $X \sim N(\mu, \sigma^2)$. *Read as X follows normal distribution with parameters μ and σ^2*.

Note:

1. The graph of $f(x)$ is famous ***bell – shaped*** curve and is symmetric about the line $X = \mu$. The top of the bell is directly above μ. For large values of σ, the curve tends

to flatten out and for small values of σ, it has a sharp peak. The curve of $f(x)$ is given below.



Normal probability curve

2. Whenever the random variable is continuous and the probabilities of it are increasing and then decreasing, in such cases we can think of using normal distribution.

Real life examples:

1) The heights of students.
2) The weights of students.
3) The diameters of bolts manufactured.
4) The lives of electrical bulbs manufactured.

3. Note that $\int_{-\infty}^{\infty} f(x)dx = 1$

Standard Normal distribution

If $X \sim N(\mu, \sigma^2)$ then $Z = \frac{X-\mu}{\sigma}$ is known as **standard normal distribution** with mean $E(Z) = 0$, with variance $V(Z) = 1$ and we write $Z \sim N(0, 1)$.Its p. d. f. is given by

$$\boldsymbol{g(z) = \frac{1}{\sqrt{2\pi}} . exp\left(-\frac{1}{2}z^2\right)}, -\infty < z < \infty$$

and its c. d. f. is given by

$$\Phi(z) = P(Z \leq z) = \int_{-\infty}^{x} g(t)\,dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} exp\left\{-\frac{1}{2}t^2\right\} dt$$

Area Property of Normal Distribution

If $X \sim N(\mu, \sigma^2)$, then $P(\mu < X < x_1) = \int_{\mu}^{x} f(x)\,dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{\mu}^{x} exp\left\{-\left(\frac{x-\mu}{\sigma}\right)^2\right\} dx$

Let $Z = \frac{X-\mu}{\sigma}$. Then $X - \mu = \sigma Z$.

If $X = \mu$, then $Z = 0$. If $X = x_1$, then $Z = \frac{X-\mu}{\sigma} = z_1$ (say).

$$\therefore P(\mu < X < x_1) = P(0 < Z < z_1) = \int_0^{z_1} g(z)\,dz = \frac{1}{\sqrt{2\pi}} \int_0^{z_1} \exp\left\{-\frac{1}{2}z^2\right\} dz$$

where $g(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$ is the p. d. f. of standard normal variate. The definite integral $\int_0^{z_1} g(z)\,dz$ is known as **normal probability integral** and gives the area under standard normal curve between the ordinates at $z = 0$ and $z = z_1$. These areas have been tabulated for different values of z_1 at intervals of 0.01 in the table given at the ***end of the module.***

In particular, the probability that the random variable X lies in the interval $(\mu - \sigma, \mu + \sigma)$ is given by

$$P(\mu - \sigma < X < \mu + \sigma) = P(-1 < Z < 1)$$

$$= \int_{-1}^{1} g(z)\,dz$$

$$= 2\int_0^{1} g(z)\,dz \qquad \text{(by symmetry)}$$

$$= 2 \times 0.3413 \qquad \text{(from table)}$$

$\Longrightarrow P(\mu - \sigma < X < \mu + \sigma) = 0.6826$

Similarly, $P(\mu - 2\sigma < X < \mu + 2\sigma) = P(-2 < Z < 2)$

$= 2 \times P(0 < Z < 2) = 2 \times 0.4772$(see table)

$\Longrightarrow P(\mu - 2\sigma < X < \mu + 2\sigma) = 0.9544$

and $P(\mu - 3\sigma < X < \mu + 3\sigma) = P(-3 < Z < 3)$

$= 2 \times P(0 < Z < 3)$

$= 2 \times 0.49865$ (see table)

$\Longrightarrow P(\mu - 3\sigma < X < \mu + 3\sigma) = 0.9973$

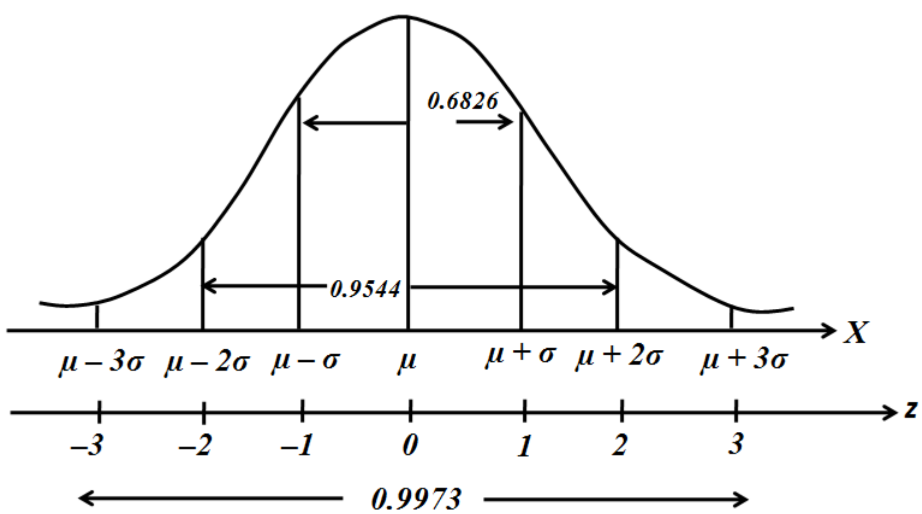
Thus, the probability that a normal variate X lies outside the range $\mu \pm 3\sigma$ is given by

$P(|x - \mu > 3\sigma|) = P(|Z| > 3) = 1 - P(|Z| \leq 3)$

$= 1 - P(-3 \leq Z \leq 3) = 1 - 0.9973 = 0.0027$

Thus, in all probability, we should expect a normal variate to lie within the range $\mu \pm 3\sigma$, though theoretically, it may range from $-\infty$ to ∞.

The probabilities computed above are exhibited in the following figure.



Note: The Gamma function defined below is used to evaluate mean and variance of the normal distribution.

$$\Gamma(n) = \int_0^\infty e^{-x} x^{n-1} dx \text{ for } n > 0$$

$$\Gamma(n+1) = n\,\Gamma(n)$$

$$\Gamma(n+1) = n!\text{, where } n \text{ is a positive integer.}$$

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

Mean of Normal distribution

$$E(X) = \int_{-\infty}^{\infty} xf(x)\,dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} xe^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

Let $z = \frac{x-\mu}{\sigma}$.Then $x = \mu + \sigma z$,$dx = \sigma dz$ and

$$E(X) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\mu + \sigma z) e^{-\frac{1}{2}z^2} dz$$

$$= \mu.\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2} dz + \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} ze^{-\frac{1}{2}z^2} dz = \mu \times 1 + 0$$

Note that the integral in first term is 1 since total probability is one and the integral in the second term in zero since the integral is an odd function.

Therefore, Mean $= E(X) = \mu$

Variance of Normal distribution

$$V(X) = E\big(X - E(X)\big)^2 = E(X-\mu)^2 \qquad (\because E(X) = \mu)$$

$$= \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-\mu)^2 e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

Let $\frac{x-\mu}{\sigma} = z$. Then $x - \mu = \sigma z$, $dx = \sigma dz$ and

$$V(X) = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2 e^{-\frac{1}{2}z^2} dz = \frac{2\sigma^2}{\sqrt{2\pi}} \int_0^{\infty} z^2 e^{-\frac{1}{2}z^2} dz$$

(Since the integrand is an even function)

Let $\frac{1}{2}z^2 = t \implies z = \sqrt{2t}$ and $dz = \frac{dt}{\sqrt{2t}}$. Then

$$V(X) = \frac{2\sigma^2}{\sqrt{2\pi}} \int_0^{\infty} 2t.e^{-t} \frac{dt}{\sqrt{2t}} = \frac{2\sigma^2}{\sqrt{\pi}} \int_0^{\infty} e^{-t}.t^{\frac{3}{2}-1} dt$$

$$= \frac{2\sigma^2}{\sqrt{\pi}} \Gamma\left(\frac{3}{2}\right)$$ (Gamma function)

$$= \frac{2\sigma^2}{\sqrt{\pi}} . \frac{1}{2} . \Gamma\left(\frac{1}{2}\right) = \frac{\sigma^2}{\sqrt{\pi}} . \sqrt{\pi} = \sigma^2$$

$$\implies V(X) = \sigma^2$$

Note: Standard deviation $= \sqrt{V(X)} = \sqrt{\sigma^2} = \sigma$

Example 2: If X is normally distributed with mean 12 and standard deviation , then

(a) Find the probabilities of the following :

(i) $X \geq 20$

(ii) $X \leq 20$ and

(iii) $0 \leq X \leq 12$

(b) Find x when $P(X > x) = 0.24$

(c) Find x_1 and x_2 when $P(x_1 < X < x_2) = 0.5$ and $P(X > x_2) = 0.25$

Solution:

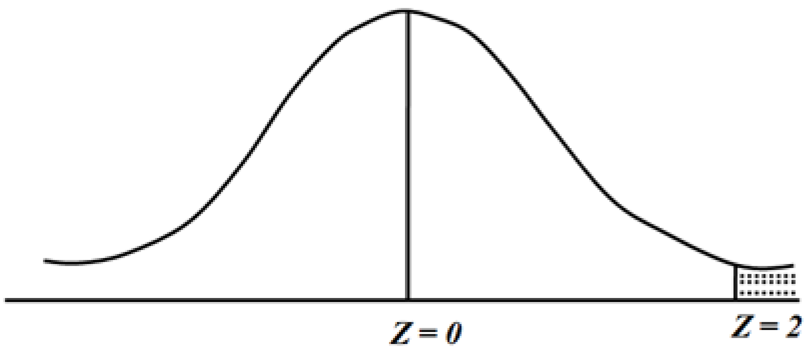
(a) it is given that $\mu = 12$ and $\sigma = 4$ $i.e.$, $X \sim N(12, 16)$

(i) Let $Z = \frac{X-12}{4}$. then $P(X \geq 20) = P\left(\frac{X-12}{4} \geq \frac{20-12}{4}\right)$

$= P(Z \geq 2) = 0.5 - P(0 \leq Z \leq 2)$

$= 0.5 - 0.4772$ (from table)

$= 0.0228$



(ii) $P(X \leq 20) = 1 - P(X \geq 20) = 1 - 0.0228 = 0.9722$

(iii) $P(0 \leq X \leq 12) = P\left(\frac{0-12}{4} \leq \frac{X-12}{4} \leq \frac{12-12}{4}\right)$

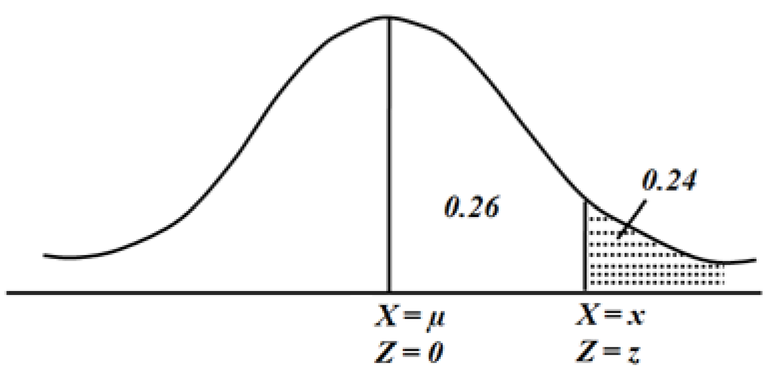
$= P(-3 \leq Z \leq 0)$

$= P(0 \leq Z \leq 3)$ (by symmetry)

$= 0.4986$ (from table)

(b) $P(X > x) = 0.24$

$\Longrightarrow P\left(\frac{X-12}{4} > \frac{x-12}{4}\right) = 0.24 \Longrightarrow P(Z > z) = 0.24$, where $z = \frac{x-12}{4}$



$\Longrightarrow P(0 < Z < z) = 0.5 - 0.24 - 0.26$

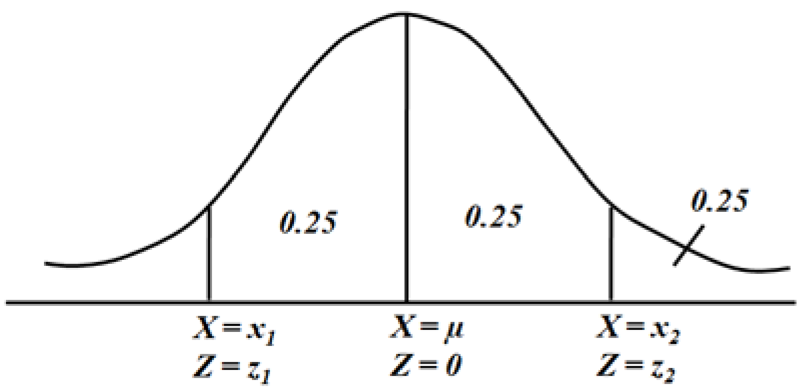
∴ From normal tables, corresponding to probability 0.26, value of $z = 0.71$(approximately)

Hence $0.71 = z = \frac{x-12}{4} \Longrightarrow x = 0.71 \times 4 + 12 = 14.84$

(c) We are given that $P(x_1 < X < x_2) = 0.5$ and $(X > x_2) = 0.25$

$\Longrightarrow P\left(\frac{x_1-12}{4} < \frac{X-12}{4} < \frac{x_2-12}{4}\right) = 0.5$ and $P\left(\frac{X-12}{4} > \frac{x_2-12}{4}\right) = 0.25$

$\Longrightarrow P(z_1 < Z < z_2) = 0.5$ and $P(Z > z_2) = 0.25$, where $z_1 = \frac{x_1-12}{4}$ and $z_2 = \frac{x_2-12}{4}$



By symmetry of normal curve, $z_1 = -z_2$. Find z_2 such that $P(0 < Z < z_2) = 0.25$

Corresponding to probability 0.25 from the normal table, we have $z_2 = 0.67$ approximately. Thus

$\frac{x_2-12}{4} = 0.67 \Longrightarrow x_2 = 12 + 4 \times 0.67 = 14.68$

Similarly, $z_1 = -z_2 \Longrightarrow \frac{x_2-12}{4} = -0.67 \Longrightarrow x_1 = 12 - 4 \times 0.67 = 9.32$

Example 3: The local authorities in a certain city install $10,000$ electric lamps in the streets of the city. If these lamps have an average life of $1,000$ burning hours with a standard deviation of 200 hours, assuming normality, what number of lamps might be expected to fail

(i) in the first 800 and 1200 burning hours?
(ii) between 800 and 1200 burning hours?

After what period of burning hours would you expect that

(a) 10% of the lamps would fail?
(b) 10% of the lamps would be still burning?

Solution:

Let X denote the life of a bulb in burning hours. Here $\mu = 1000,\ \sigma = 200$ and $X \sim N(1000, 40000)$

(i) Find $P(X < 800) = P\left(\frac{X-1000}{200} < \frac{800-1000}{200}\right)$

$= P(Z < -1)$, where $Z = \frac{X-1000}{200} \sim N(0,1)$

$= P(Z > 1) = 0.5 - P(0 < Z < 1)$

$= 0.5 - 0.3413 = 0.1587$

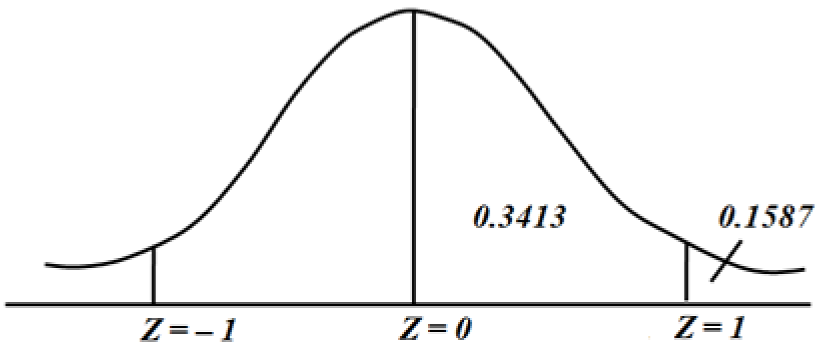
$\therefore$ Out of $10{,}000$ bulbs, number of bulbs which fail in the first 800 hours is

$10{,}000 \times 0.1587 = 1{,}587.$

(ii) Find $P(800 < X < 1200) = P\left(\frac{800-1000}{200} < \frac{X-1000}{200} < \frac{1200-1000}{200}\right)$

$= P(-1 < Z < 1) = 2.P(0 < Z < 1)$

$= 2 \times 0.3413 = 0.6826$



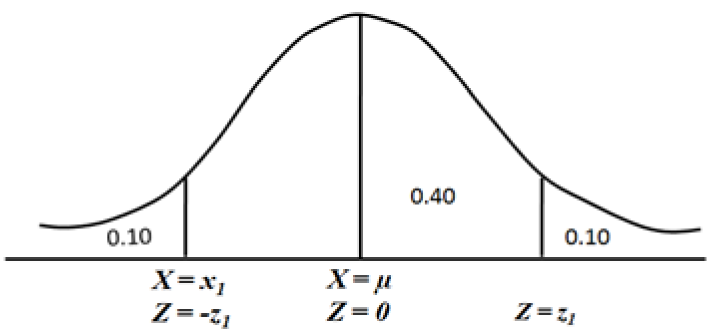
Hence, the expected number of bulbs with life between 800 and 1200 hours of burning life is $10{,}000 \times 0.6826 = 6{,}826$.

(a) Let 10% of the bulbs fail after x_1 hours of burning life. Then we have to find x_1 such that

$$P(X < x_1) = 0.10 \Rightarrow P\left(\frac{X-1000}{200} < \frac{x_1-1000}{200}\right) = 0.10$$
$$\Rightarrow P(Z < -z_1) = 0.10 \text{ , where } z_1 = -\left(\frac{x_1-1000}{200}\right)$$
$$\Rightarrow P(Z > z_1) = 0.10$$
$$\Rightarrow P(0 < Z < z_1) = 0.5 - 0.10 = 0.40$$

From table corresponding to probability 0.40, we have

$$z_1 = 1.28 \Rightarrow -\left(\frac{x_1-1000}{200}\right) = 1.28$$
$$\Rightarrow x_1 = 1000 - 1.28 \times 200 = 1000 - 256 = 744.$$



Thus, after 744 hours of burning life, 10% of the bulbs will fail.

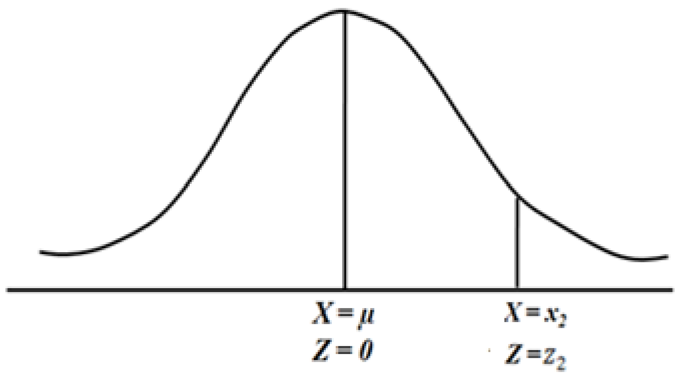
(b) Let 10% of the bulbs be still burning after x_2 hours of burning life. Then we have

$$P(X > x_2) = 0.10 \Rightarrow P\left(\frac{X-1000}{200} > \frac{x_2-1000}{200}\right) = 0.10$$

$\Rightarrow P(Z > z_2) = 0.10$,where $z_2 = \frac{x_2-1000}{200}$

From normal tables, $z_2 = 1.28$ and hence

$\frac{x_2-1000}{200} = 1.28 \Rightarrow x_2 = 1000 + 1.28 \times 200 = 1000 + 256 = 1256$



Hence, after 1,256 hours of burning life, 10% of the bulbs will be still burning.

De Moivre-Laplace Theorem (Normal Approximation to Binomial Distribution)

Let $X \sim B(n, p)$. Then its p.m.f. is given by $p(x) = \binom{n}{x} p^x q^{n-x}$ for $x = 0,1,2, \dots, n$. The mean and variance of X are given by $\mu = np$ and $\sigma^2 = npq$ respectively. Now, $P(k_1 \le X \le k_2) = \sum_{x=k_1}^{k_2} \binom{n}{x} p^x q^{n-k}$ for some non-negative integers k_1 and k_2 such that $k_1 < k_2$. Since the binomial coefficient $\binom{n}{x}$ grows quite rapidly with n, it is very difficult to compute $P(k_1 \le X \le k_2)$ for large n. In this context, normal approximation to binomial distribution is extremely useful.

Let $Z = \frac{X-\mu}{\sigma} = \frac{X-np}{\sqrt{npq}}$. If n is large with neither p nor q close to zero, the binomial distribution can be approximated by the standard normal distribution. Thus,

$$\lim_{n\to\infty} P(k_1 \le X \le k_2) = \lim_{n\to\infty} P\left(\frac{k_1-np}{\sqrt{npq}} \le Z \le \frac{k_2-np}{\sqrt{npq}}\right) = \frac{1}{\sqrt{2\pi}} \int_{z_1}^{z_2} e^{-\frac{1}{2}z^2} dz$$

$$where\ z_1 = \frac{k_1-np}{\sqrt{npq}}\ and\ z_2 = \frac{k_2-np}{\sqrt{npq}}$$

This is a very good approximation when both np and npq are greater than 5.

Example 4: A coin is tossed 10 times. Find the probability of getting between 4 and 7 heads inclusive using the (a) binomial distribution and (b) the normal approximation to the binomial distribution.

Solution:

(a) Let X denote the number of heads in 10 tosses. Then $X \sim B\left(10, \frac{1}{2}\right)$ and $\mu = np = 5$ and $\sigma^2 = npq = 2.5$ and

$$P(4 \le X \le 7) = \sum_{x=4}^{7} p(x) = \sum_{x=4}^{7} \binom{n}{x} \left(\frac{1}{2}\right)^{x} \left(\frac{1}{2}\right)^{10-x} = \sum_{x=4}^{7} \binom{n}{x} \left(\frac{1}{2}\right)^{10}$$

$$= \frac{\binom{10}{4} + \binom{10}{5} + \binom{10}{6} + \binom{10}{7}}{1024} = \frac{792}{1024} = 0.7734$$

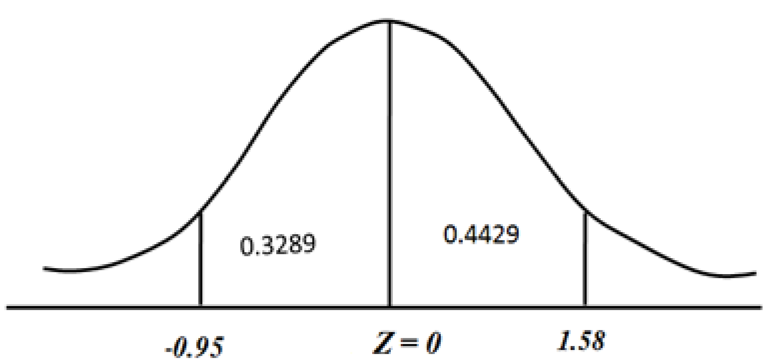
(b) The discrete binomial probability distribution is approximated to continuous normal probability distribution. The integers 4,5,6,7 lie in the interval (3.5 to 7.5). Thus,

$$P(4 \le X \le 7) = P(3.5 \le X \le 7.5) = P\left(\frac{3.5-5}{\sqrt{25}} \le Z \le \frac{7.5-5}{\sqrt{2.5}}\right)$$

$$= P(-0.95 \le Z \le 1.58) = P(-0.95 \le Z \le 0) + P(0 \le Z \le 1.58)$$

$$= P(0 \le Z \le 0.95) + P(0 \le Z \le 1.58)$$

$$= 0.3289 + 0.4429 = 0.7718$$



Exponential distribution: A c.r.v. X is said to follow **exponential distribution** with parameter λ if its p.m.f. is given by

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

The c.d.f. is given by

$$F(x) = P(X \leq x) = \int_0^x f(t)\,dt = \lambda \int_0^x e^{-\lambda t} dt = \lambda \left[\frac{e^{-\lambda t}}{-\lambda} \right]_0^x = 1 - e^{-\lambda x}$$

$$\Rightarrow F(x) = 1 - e^{-\lambda x}$$

Notation: $X \sim E(\lambda)$. Read as X *follows exponential distribution with parameter* λ.

Real life examples of exponential distribution

1. The time taken to serve a customer at a petrol pump, railway booking counter or any other service facility.
2. The period of time for which an electronic component operates without any breakdown.
3. The time between two successive arrivals at any service facility.

Mean and Variance of exponential distribution

For $r \geq 1$, $E(X^r) = \int_0^\infty x^r f(x)\,dx = \lambda \int_0^\infty x^r e^{-\lambda x} dx$

Let $\lambda x = t$. Then t varies between 0 to ∞ and $dx = \frac{dt}{\lambda}$. Then

$$E(X^r) = \lambda \int_0^\infty \left(\frac{t}{\lambda} \right)^r e^{-t} \frac{dt}{\lambda} = \frac{1}{\lambda^r} \int_0^\infty e^{-t} t^{(r+1)-1} dt$$

$$\Rightarrow E(X^r) = \frac{\Gamma(r+1)}{\lambda^r} = \frac{r!}{\lambda^r} \text{(using Gamma function)}$$

Thus, mean $= \mu = E(X) = \frac{1}{\lambda}$ and $E(X^2) = \frac{2!}{\lambda^2} = \frac{2}{\lambda^2}$

Hence $\sigma^2 = V(X) = E(X^2) - \left(E(X)\right)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$

Therefore, $\mu = \frac{1}{\lambda}$ and $\sigma^2 = \frac{1}{\lambda^2}$.

Example 5: Assume that the length of phone calls made at a particular telephone booth is exponentially distributed with a mean of 3 minutes. If you arrive at the telephone booth just as Ramu was about to make a call, find the following:

a. The probability that you will wait more than 5 minutes before Ramu is done with the call.
b. The probability that Ramu's call will last between 2 minutes and 6 minutes.

Solution: Let X be a r.v. that denotes the length of calls made at the telephone booth. Since the mean length of calls $\frac{1}{\lambda} = 3$, the p.d.f. is given by

$$f(x) = \frac{1}{3} e^{-\frac{x}{3}}$$

a. $P(X > 5) = \int_5^\infty f(x)dx = \frac{1}{3}\int_5^\infty e^{-\frac{x}{3}}dx = \left[e^{-\frac{x}{3}}\right]_5^\infty = e^{-\frac{5}{3}}$

b. $P[2 \le X \le 6] = \int_2^6 f(x)dx = \frac{1}{3}\int_2^6 e^{-\frac{x}{3}}dx = \left[-e^{-\frac{x}{3}}\right]_2^6 = e^{-\frac{2}{3}} - e^{-2}$

Memory lessness property of exponential distribution

The exponential distribution is used extensively in reliability engineering to model the lifetimes of systems. Suppose the life X of an equipment is exponentially distributed with a mean of $\frac{1}{\lambda}$. Assume that the equipment has not failed by time t. We want to find the probability that $X \le t + s$ given that $X > t$ for some nonnegative additional time s.

Thus,

$$P(X \le s + t | X > t) = \frac{P(X \le s+t, X > t)}{P(X > t)} = \frac{P(t < X \le s+t)}{P(X>t)} = \frac{F(s+t) - F(t)}{1 - F(t)}$$

$$= \frac{(1 - e^{-\lambda(s+t)}) - (1 - e^{-\lambda t})}{e^{-\lambda t}} = \frac{e^{-\lambda t} - e^{-\lambda(s+t)}}{e^{-\lambda t}} = 1 - e^{-\lambda s} = F(s) = P(X \le s)$$

$$\Rightarrow P(X \le s + t | X > t) = P(X \le s)$$

This indicates that the process only remembers the present and not the past.

Example 6: In example 5, Ramu, who is using the phone at the telephone booth, had already talked for 2 minutes before you arrived. According to the memory lessness property of the exponential distribution, the mean time until Ramu is done with the call is still 3 minutes. The random variable forgets the length of time the call had lasted before you arrived.

Relationship between exponential and Poisson distributions

Let λ denote the mean number of arrivals per unit of time, say per second. Then the mean number of arrivals in t seconds is λt.

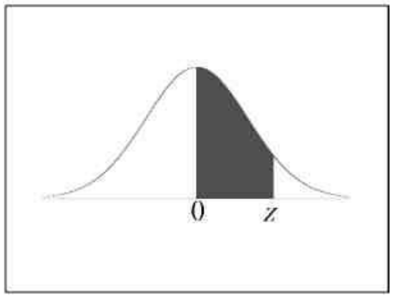
Let X denote the number of arrivals during an interval of t seconds.

Let Y denote the time between two successive arrivals.

If $X \sim P(\lambda t)$ i.e., $p(x) = P(X = x) = \frac{e^{-\lambda t}(\lambda t)^x}{x!}$ for $x = 0,1,2, \ldots; \quad t \geq 0$, then

$Y \sim E(\lambda)$ i.e., $f(x) = \lambda e^{-\lambda t}$.

Standard Normal Distribution Table



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	. 3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	. 3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4 332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952

2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998
3.5	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998

2.6

Functions of Random Variables

The previous modules discussed basic properties of events defined in a given sample space and the random variables used to represent those events. The fundamental assumption that was made in those modules is that events can always be defined by **random variables**. However, in many applications, the events are functions of other events. For example, the time until a complex system fails is a function of the time to failure of the individual components that make up the system. This means that the random variable used to represent the time to failure of the complex system is a function of the random variables used to represent the times to failure of the component parts of the system. This module deals with functions of random variables. Because of the complexity involved in computing the c.d.fs and p.d.fs of multiple random variables, the discussion is restricted to functions of at most two random variables.

Functions of One Random Variable: Let X be a r.v. with p.d.f. (or p.m.f.) $f_X(x)$ and c.d.f. $F_X(x)$. Let Y be the new random variable that is a function of X. That is,

$$Y = g(X)$$

Then we are interested in computing p.d.f (or p.m.f.) $f_Y(y)$ and c.d.f. $F_Y(y)$ of Y.

For example, let $Y = X + 5$. Then

$$F_Y(y) = P(Y \leq y) = P[X + 5 \leq y] = P[X \leq y - 5] = F_X(y - 5)$$

Linear Functions: Consider the function $g(X) = aX + b$, where a and b are constants. The c.d.f of Y is given by

$$F_Y(y) = P\{Y \leq y\} = P[aX + b \leq y]$$

$$= P\left[X \leq \frac{y-b}{a}\right] = F_X\left(\frac{y-b}{a}\right)$$

where a is positive .The p.d.f. of Y is given by

$$f_Y(y) = \frac{d}{dy}\left(F_Y(y)\right) = \frac{d}{dy}\left(F_X\left(\frac{y-b}{a}\right)\right) = \left(\frac{d}{du}\left(F_X(u)\right)\right)\left(\frac{du}{dy}\right)$$

where $u = \frac{y-b}{a}$ and $\frac{du}{dy} = \frac{1}{a}$. Thus,

$$f_Y(y) = \left(\frac{1}{a}\right) f_X(u) = \left(\frac{1}{a}\right) f_X\left(\frac{y-b}{a}\right)$$

If $a < 0$, we have,

$$F_Y(y) = P(Y \leq y) = P(aX + b \leq y) = P(aX \leq y - b)$$

$$= P\left(X \geq \frac{y-b}{a}\right) = 1 - \left\{P\left[X \leq \frac{y-b}{a}\right] - P\left[X = \frac{y-b}{a}\right]\right\} \quad (\because a < 0)$$

The change in sign on the second line arises from the fact that a is negative. If X is continuous, $P\left[X = \frac{(y-b)}{a}\right] = 0$. Thus, the c.d.f and p.d.f for the case of negative a are given by

$$F_Y(y) = 1 - P\left[X \leq \frac{y-b}{a}\right]$$

$$= 1 - F_X\left(\frac{y-b}{a}\right)$$

Therefore, $f_Y(y) = \frac{d}{dy}\left(F_Y(y)\right) = -\left(\frac{1}{a}\right) f_X\left(\frac{y-b}{a}\right)$

Therefore, the general p.d.f. of Y is given by

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$$

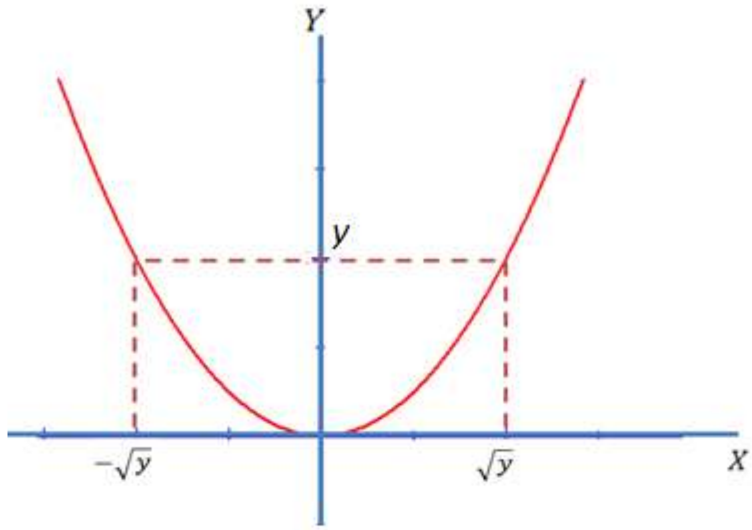
Example 1: Find the p.d.f of Y in terms of the p.d.f of X if $Y = 2X + 7$.

Solution: From the results obtained above,

$$F_Y(y) = F_X\left(\frac{y-7}{2}\right)$$

and $$f_Y(y) = \left(\frac{1}{2}\right) f_X\left(\frac{y-7}{2}\right)$$

Power Functions: Consider the quadratic function $Y = X^2$. The plot of Y against X is shown in the following figure where we see that for one value of Y there are two values of X, namely $\sqrt{Y}$ and $-\sqrt{Y}$.



Thus, the c.d.f of Y is given by

$$F_Y(y) = P[Y \leq y] = P[X^2 \leq y]$$

$$= P[|X| \leq \sqrt{y}], \quad y > 0$$

$$= P[-\sqrt{y} \leq X \leq \sqrt{y}]$$

$$= F_X(\sqrt{y}) - F_X(-\sqrt{y})$$

The p.d.f of Y is given by

$$f_Y(y) = \frac{dF_Y(y)}{dy} = \frac{d}{dy}\left[F_X(\sqrt{y}) - F_X(-\sqrt{y})\right]$$

Let $u = \sqrt{y} = y^{\frac{1}{2}}$. Thus, $\frac{du}{dy} = \frac{1}{2}y^{-\frac{1}{2}}$ and

$$f_Y(y) = \frac{d}{dy}\left[F_X(\sqrt{y}) - F_X(-\sqrt{y})\right]$$

$$= \frac{d}{du}\left(F_X(u)\right)\frac{du}{dy} + \frac{d}{du}\left(F_X(-u)\right)\frac{du}{dy}$$

$$=\frac{1}{2}y^{-\frac{1}{2}}\left[\frac{d}{du}\big(F_X(u)\big)+\frac{d}{du}\big(F_X(-u)\big)\right]$$

$$=\frac{1}{2}y^{-\frac{1}{2}}\left[f_X\left(\sqrt{y}\right)+f_X\left(-\sqrt{y}\right)\right]$$

$$=\frac{f_X(\sqrt{y})+f_X(-\sqrt{y})}{2\sqrt{y}}\quad ,y>0$$

If $f_X(x)$ is an even function, then $f_X(x)=f_X(-x)$ and $F_X(-x)=1-F_X(x)$. Thus, we have

$$f_Y(y)=\frac{f_X(\sqrt{y})+f_X(-\sqrt{y})}{2\sqrt{y}}=\frac{2f_X(\sqrt{y})}{2\sqrt{y}}=\frac{f_X(\sqrt{y})}{\sqrt{y}}$$

Example2: Find the p.d.f of the random variable $Y=X^2$, where X is the standard normal random variable.

Solution: Since the p.d.f. of X is given by $f_X(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$, which is an even function, we know that

$$F_Y(y)=F_X\left(\sqrt{y}\right)-F_X\left(-\sqrt{y}\right)$$

$$=2F_X\left(\sqrt{y}\right)-1$$

Therefore, if we let $u=\sqrt{y}$, then

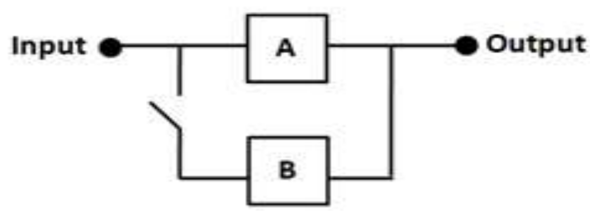
$$f_Y(y)=\frac{dF_Y(y)}{dy}=2\frac{dF_X(u)}{du}\frac{1}{2\sqrt{y}}$$

$$=\frac{1}{\sqrt{y}}f_X\left(\sqrt{y}\right)$$

$$=\frac{1}{\sqrt{2\pi y}}e^{-\frac{y}{2}},\ y>0$$

Sum of Two Independent Random variables

Consider two independent continuous random variables X and Y. We are interested in computing the c.d.f and p.d.f of their sum $g(X,Y) = S = X + Y$. The random variable S can be used to model the reliability of systems with stand-by connections, as shown in *fig*.1 In such systems, the component A whose time-to-failure is represented by the random variable X is the primary component, and the component B whose time-to-failure is represented by the random variable Y is the backup component that is brought into operation when the primary component fails. Thus, S represents the time until the system fails, which is the sum of the lifetimes of both components.



$fig.1$

Their c.d.f. can be obtained as follows:

$$F_S(s) = P[S \le s] = P[X + Y \le s] = \iint_D f_{XY}(x,y)\,dxdy$$

where D is the set $D = \{(x,y) | x + y \le s\}$, which is the area to the left of the line $s = x + y$ as shown in *fig.* 2.

Thus,

$$F_S(s) = \int_{-\infty}^{\infty}\int_{-\infty}^{s-y} f_{XY}(x,y)\,dxdy = \int_{-\infty}^{\infty}\int_{-\infty}^{s-y} f_X(x)\,f_Y(y)\,dxdy$$

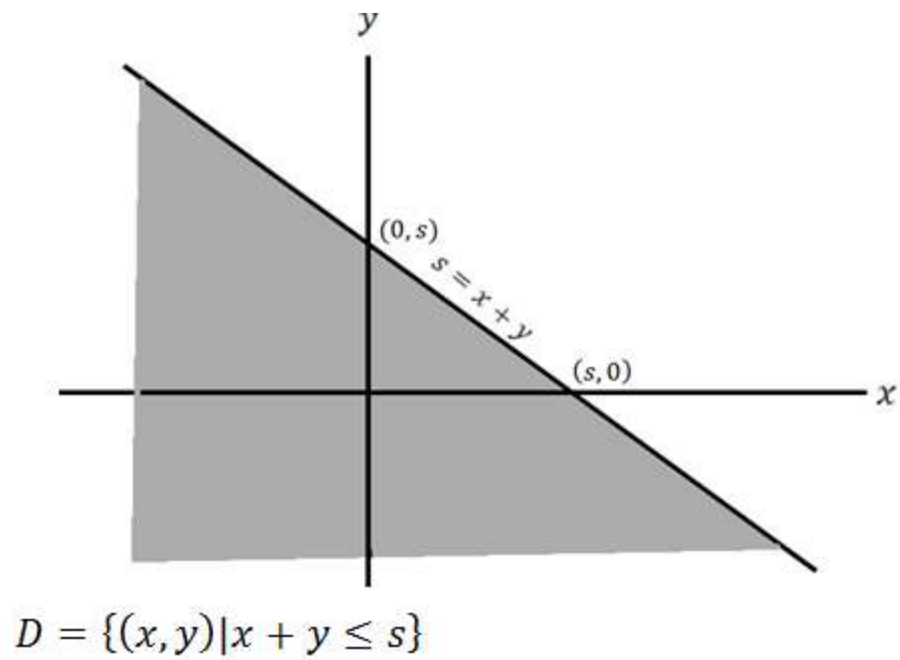
$$= \int_{-\infty}^{\infty}\left\{\int_{-\infty}^{s-y} f_X(x)\,dx\right\} f_Y(y)\,dy$$

$$= \int_{-\infty}^{\infty} F_X(s-y)\,f_Y(y)\,dy$$

The p.d.f. of S is obtained by differentiating the c.d.f. , as follows:

$$f_S(s) = \frac{d}{ds} F_S(s) = \frac{d}{ds} \int_{-\infty}^{\infty} F_X(s-y) f_Y(y) dy$$
$$= \int_{-\infty}^{\infty} \frac{d}{ds} F_X(s-y) f_Y(y) dy$$
$$= \int_{-\infty}^{\infty} f_X(s-y) f_Y(y) dy$$

where we have assumed that we can interchange differentiation and integration. The expression on the right-hand side is a well-known result in signal analysis



$fig. 2$

called the **convolution integral**. Thus, we find that the p.d.f of the sum S of two independent random variables X and Y is the convolution of the p.d.fs of the two random variables; that is,

$$f_S(s) = f_X(s) f_Y(s)$$

Example 3: Find the p.d.f. of the sum of X and Y if the two random variables are independent random variables with the common p.d.f.

$$\boldsymbol{f_X(u) = f_Y(u)} = \begin{cases} \mathbf{\frac{1}{4}} & \mathbf{0} < u < 4 \\ \mathbf{0} & \textbf{otherwise} \end{cases}$$

Solution: The limits of integration of the p.d.f of $S = X + Y$ can be computed with the aid of $fig.3$ When $0 \le s \le 4$ (see $fig.3\ (a)$ where $f_Y(s-x)$ is shown in dashed lines),

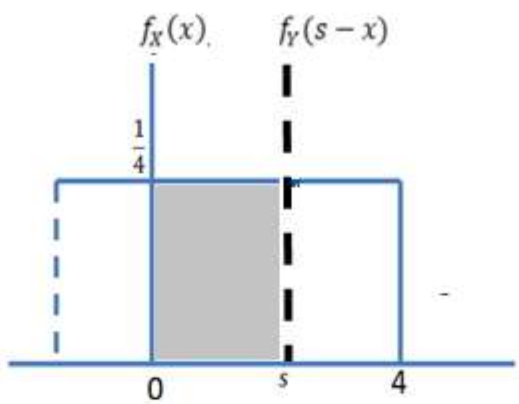
$$f_S(s) = \int_0^s \frac{1}{16}\,dy = \frac{s}{16}$$

For $4 < s < 8$ (see $fig.3\ (b)$), we obtain

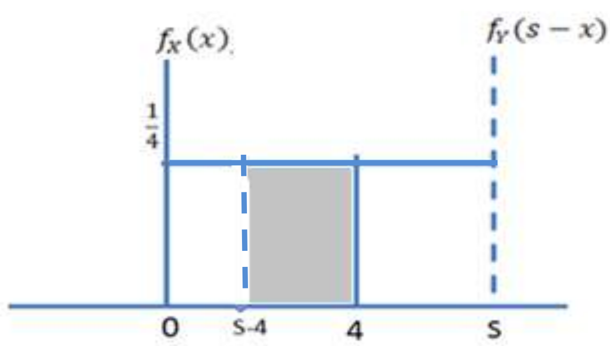
$$f_S(s) = \int_{s-4}^4 dy = \frac{8-s}{16}$$

Thus ,

$$f_S(s) = \begin{cases} \frac{s}{16} & ,\ 0 \le s \le 4 \\ \frac{8-s}{16} & ,\ 4 < s < 8 \\ 0 & ,\ \text{otherwise} \end{cases}$$



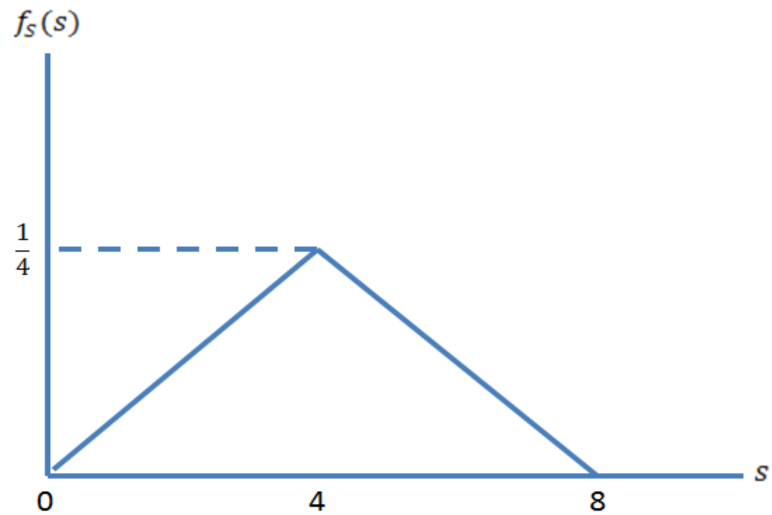
$fig\ 3(a)$



$fig\ 3(b)$

$Fig.3$: Convolution of p.d.fs $(a)\ \ 0 \le s \le 4$ and $(b)\quad 4 \le s \le 8$

The p.d.f of $S = X + Y$ is illustrated in the following figure.



Example 4: The time X between consecutive snowstorms in winter is a random variable with the p.d.f.

$$\boldsymbol{f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & \text{otherwise} \end{cases}}$$

Assume it has not snowed up until now. What is the p.d.f. of the time U until the second snowstorm?

Solution: Let X be the random variable that denotes the time until the first snowstorm from the reference time, and let Y be the random variable that denotes the time between the first snowstorm and the second snowstorm. If we assume that the times between snowstorms are independent, then X and Y are independent and identically distributed random variables. That is, the p.d.f of Y is given by

$$f_Y(y) = \begin{cases} \lambda e^{-\lambda y}, & y \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

Thus, $U = X + Y$, and the p.d.f. of U is given by

$$f_U(u) = \int_0^{\infty} f_X(x) f_Y(u - x) dx$$

Since $f_X(x) = 0$ when $x < 0, f_Y(u-x) = 0$ when $u - x < 0$ (or $x > u$). Thus, the range of interest in the integration is $0 \le x \le u$, and we obtain

$$\begin{aligned} f_U(u) &= \int_0^u f_X(x) f_Y(u-x)dx \\ &= \int_0^u \lambda e^{-\lambda x} \lambda e^{-\lambda(u-x)} dx = \lambda^2 e^{-\lambda u} \int_0^u dx \\ &= \lambda^2 u e^{-\lambda u} \quad u \ge 0 \end{aligned}$$

This is the Erlang – 2 distribution.

Note: A random variable X is said to follow Erlang-k distribution if its p.d.f. is given by

$$f(x) = \begin{cases} \frac{\lambda^k x^{k-1} e^{-\lambda x}}{(k-1)!} & , \quad k = 1,2,\dots;\ \lambda > 0; x \ge 0 \\ 0 & , \quad x < 0 \end{cases}.$$

Sum of Two Discrete Random Variables

The examples above deal with continuous random variables.
Let $Z = X + Y$, where X and Y are discrete random variables. Then the p.m.f of Z is given by

$$\begin{aligned} p_Z(z) = P[Z = z] = P[X + Y = z] &= \sum_{k \le z} P[X = k, Y = z - k] \\ &= \sum_{k \le z} p_{XY}[k, z - k] \end{aligned}$$

If X and Y are independent random variables, then the p.m.f. of Z is the convolution of the p.m.f of X and the p.m.f of Y. That is,

$$p_Z(z) = \sum_{k \le z} p_{XY}(k, z - k) = \sum_{k \le z} p_X(k) p_Y(z - k)$$

Sum of Two Independent Binomial Random Variables

Let X and Y be two independent binomial random variables with parameters (n, p) and (m, p), respectively and their sum be Z; that is, $Z = X + Y$. Then the p.m.f of Z is given by

$$
\begin{aligned}
p_Z(z) &= P[X + Y = z] \\
&= \sum_{k=0}^{n} P[X = k, Y = z - k] = \sum_{k=0}^{n} P[X = k] P[Y = z - k] \\
&= \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \binom{m}{z-k} p^{z-k} (1-p)^{m-z+k} \\
&= p^z (1-p)^{n+m-z} \sum_{k=0}^{n} \binom{n}{k} \binom{m}{z-k}
\end{aligned}
$$

Using the combinatorial identity $\binom{n+m}{z} = \sum_{k=0}^{n} \binom{n}{k}\binom{m}{z-k}$, we obtain

$$p_Z(z) = \binom{n+m}{z} p^z (1-p)^{n+m-z}$$

This result shows that the sum of two independent binomial random variables with parameters (n, p) and (m, p) is a binomial random variable with parameter $(n + m, p)$.

Minimum of Two Independent Random Variables

Consider two independent continuous random variables X and Y. We are interested in a random variable U that is the minimum of X and Y; that is, $U = \min(X, Y)$. The random variable U can be used to represent the reliability of systems with series connections, as shown in $fig. 4$ .Such systems are operational as long as all components are operational. The first component to fail causes the system to fail. Thus, if in the example shown in $fig. 4$, the times-to-failure are represented by the random variables X and Y, then S represents the time until the system fails, which is the minimum of the lifetimes of the two components.

The c.d.f. of U can be obtained as follows:

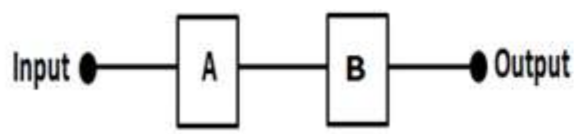
$$F_U(u) = P[U \leq u] = P[\min(X, Y) \leq u] = P[(X \leq u, X \leq Y) \cup (Y \leq u, X > Y)]$$

Since $P[A \cup B] = P[A] + P[B] - P[A \cap B]$, we have that $F_U(u) = F_X(u) + F_Y(u) - F_{XY}(u, u)$. Also, since X and Y are independent, we obtain the c.d.f. and p.d.f. of U as follows:

$$F_U(u) = F_X(u) + F_Y(u) - F_{XY}(u, u) = F_X(u) + F_Y(u) - F_X(u)F_Y(u)$$

$$f_U(u) = \frac{d}{du}F_U(u) = f_X(u) + f_Y(u) - f_X(u)F_Y(u) - F_X(u)f_Y(u)$$

$$= f_X(u)\{1 - F_Y(u)\} + f_Y(u)\{1 - F_X(u)\}$$



$fig. 4$

$fig. 4$: Series connection modeled by random variable U

Example 5: Assume that $U = \min(X, Y)$, where X and Y are independent random variables with the respective p.d.fs

$$\boldsymbol{f_X(x) = \lambda e^{-\lambda x} \quad x \geq 0}$$

$$\boldsymbol{f_Y(y) = \mu e^{-\mu y} \quad y \geq 0}$$

where $\lambda > 0$ and $\mu > 0$. What is the p.d.f. of U?

Solution: We first obtain the c.d.fs of X and Y, which are as follows:

$$F_X(x) = P[X \leq x] = \int_0^x \lambda e^{-\lambda w}\, dw = 1 - e^{-\lambda x}$$

$$F_Y(y) = P[Y \leq y] = \int_0^y \mu e^{-\mu w}\, dw = 1 - e^{-\mu y}$$

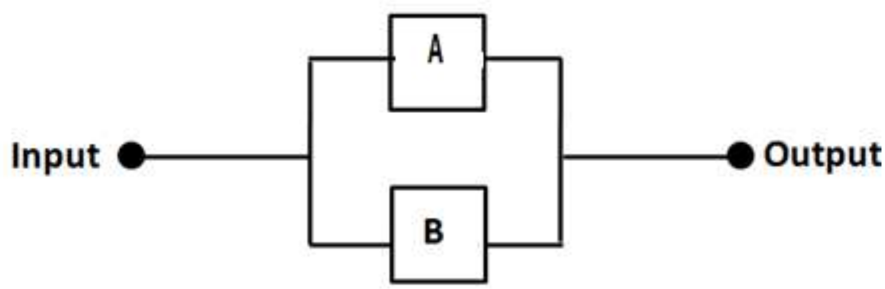
Thus, the p.d.f of U is given by

$$f_U(u) = f_X(u)\{1 - F_Y(u)\} + f_Y(u)\{1 - F_X(u)\}$$
$$= \lambda e^{-\lambda u} e^{-\mu u} + \mu e^{-\mu u} e^{-\lambda u}$$
$$= (\lambda + \mu) e^{-(\lambda+\mu)u} \ , \ u \geq 0$$

This is exponential distribution with mean $\frac{1}{\lambda+\mu}$.

Maximum of Two Independent Random Variables

Consider two independent continuous random variables X and Y. We are interested in the c.d.f. and p.d.f. of the random variable W that is the maximum of the two random variables; that is, $W = \max(X, Y)$. The random variable W can be used to represent the reliability of systems with parallel connections, as shown in



$fig.5$

$fig.5$: Parallel connection modeled by the random variable W

In such systems, we are interested in passing a signal between the two endpoints through either the component labeled A or the component labeled B. Thus, as long as one or both components are operational, the system is operational. This implies that the system is declared to have failed when both paths become unavailable. That is, the reliability of the system depends on the reliability of the last component to fail.

The c.d.f of W can be obtained by noting that if the greater of the two random variables is less than or equal to w, then the smaller random variable must also be less than or equal to w. Thus,

$$F_W(w) = P[W \leq w] = P[\max(X,Y) \leq w] = P[(X \leq w) \cap (Y \leq w)]$$
$$= F_{XY}(w,w)$$

Since X and Y are independent, we obtain the c.d.f and p.d.f of W as follows:

$$F_W(w) = F_{XY}(w,w) = F_X(w)F_Y(w)$$

$$f_W(w) = \frac{d}{dw}F_W(w) = f_X(w)F_Y(w) + F_X(w)f_Y(w)$$

Example 6: Assume that $W = \max(X,Y)$, where X and Y are independent random variables with the respective p.d.fs:

$$\boldsymbol{f_X(x) = \lambda e^{-\lambda x} \quad x \geq 0}$$

$$\boldsymbol{f_Y(y) = \mu e^{-\mu y} \quad y \geq 0}$$

where $\lambda > 0$ and $\mu > 0$. What is the pdf of W.

Solution: We first obtain the c.d.fs of X and Y, which are as follows:

$$F_X(x) = P[X \leq x] = \int_0^x \lambda e^{-\lambda z}\, dz = 1 - e^{-\lambda x}$$

$$F_Y(y) = P[Y \leq y] = \int_0^y \mu e^{-\mu z}\, dz = 1 - e^{-\mu y}$$

Thus, the p.d.f of W is given by

$$f_W(w) = f_X(w)F_Y(w) + F_X(w)f_Y(w)$$
$$= \lambda e^{-\lambda w}(1 - e^{-\mu w}) + \mu e^{-\mu w}\left(1 - e^{-\lambda w}\right)$$
$$= \lambda e^{-\lambda w} + \mu e^{-\mu w} - (\lambda + \mu)e^{-(\lambda+\mu)w} \quad w \geq 0$$

Note that the mean of W is $\frac{1}{\lambda} + \frac{1}{\mu} - \frac{1}{\lambda+\mu}$.

Two Functions of Two Random Variables

Let X and Y be two random variables with a given joint p.d.f $f_{XY}(x,y)$. Assume that U and W are two functions of X and Y; that is, $U = g(X,Y)$ and $W = h(X,Y)$. Sometimes it is necessary to obtain the joint p.d.f of U and W, $f_{UW}(u,w)$, in terms of the p.d.fs of X and Y.

It can be shown that $f_{UW}(u,w)$ is given by

$$f_{UW}(u,w) = \frac{f_{XY}(x_1,y_1)}{|J(x_1,y_1)|} + \frac{f_{XY}(x_2,y_2)}{|J(x_2,y_2)|} + \cdots + \frac{f_{XY}(x_n,y_n)}{|J(x_n,y_n)|}$$

where $(x_1,y_1),(x_2,y_2),\ldots,(x_n,y_n)$ are real solutions of the equations $u = g(x,y)$ and $w = h(x,y)$; and $J(x,y)$ is called the **Jacobian** of the transformation $\{u = g(x,y), w = h(x,y)\}$ and defined by

$$J(x,y) = \begin{vmatrix} \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \\ \frac{\partial h}{\partial x} & \frac{\partial h}{\partial y} \end{vmatrix} = \frac{\partial g}{\partial x}\cdot\frac{\partial h}{\partial y} - \frac{\partial g}{\partial y}\cdot\frac{\partial h}{\partial x}$$

Example 7: Let $\boldsymbol{U = g(X,Y) = X + Y}$ and $\boldsymbol{W = h(X,Y) = X - Y}$. Find $\boldsymbol{f_{UW}(u,w)}$.

Solution: The unique solution to the equations $u = x + y$ and $w = x - y$ is $x = \frac{u+w}{2}$ and $y = \frac{u-w}{2}$. Thus, there is only one set of solutions. Since

$$J(x,y) = \begin{vmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial w}{\partial x} & \frac{\partial w}{\partial y} \end{vmatrix} = \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} = -2$$

we obtain

$$f_{UW}(u,w) = \frac{f_{XY}(x,y)}{|J(x,y)|} = \frac{1}{|-2|} f_{XY}\left(\frac{u+w}{2},\frac{u-w}{2}\right) = \frac{1}{2} f_{XY}\left(\frac{u+w}{2},\frac{u-w}{2}\right)$$

Application of the Transformation Method

Assume that $U = g(X, Y)$, and we are required to find the p.d.f. of U. We can use the above transformation method by defining an auxiliary function $W = X$ or $W = Y$ so we can obtain the joint p.d.f. $f_{UW}(u, w)$ of U and W. Then we obtain the required marginal p.d.f. $f_U(u)$ as follows:

$$f_U(u) = \int_{-\infty}^{\infty} f_{UW}(u, w)dw$$

Example 8: Find the p.d.f. of the random variable $\boldsymbol{U = X + Y}$, where the joint p.d.f. of $\boldsymbol{X}$ and $\boldsymbol{Y}$, $\boldsymbol{f_{XY}(x, y)}$, is given.

Solution: We define the auxiliary random variable $W = X$. Then the unique solution to the two equations is $x = w$ and $y = u - w$, and the Jacobian of the transformation is

$$J(x, y) = \begin{vmatrix} \dfrac{\partial u}{\partial x} & \dfrac{\partial u}{\partial y} \\ \dfrac{\partial w}{\partial x} & \dfrac{\partial w}{\partial y} \end{vmatrix} = \begin{vmatrix} 1 & 1 \\ 1 & 0 \end{vmatrix} = -1$$

Since there is only one solution to the equations, we have that

$$f_{UW}(u, w) = \frac{f_{XY}(w, u - w)}{|-1|} = f_{XY}(w, u - w)$$

$$f_U(u) = \int_{-\infty}^{\infty} f_{UW}(u, w)dw = \int_{-\infty}^{\infty} f_{XY}(w, u - w)dw$$

This reduces to the convolution integral. We obtained earlier when X and Y are independent.

Example 9: Find the p.d.f. of the random variable $U = X - Y$, where the joint p.d.f. of X and Y is given.

Solution: We define the auxiliary random variable $W = X$. Then the unique solution to the two equations is $x = w$ and $y = w - u$, and the Jacobian of the transformation is

$$J(x,y) = \begin{vmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial w}{\partial x} & \frac{\partial w}{\partial y} \end{vmatrix} = \begin{vmatrix} 1 & -1 \\ 1 & 0 \end{vmatrix} = 1$$

Since there is only one solution to the equations, we have that

$$f_{UW}(u,w) = f_{XY}(w, w-u)$$

$$f_U(u) = \int_{-\infty}^{\infty} f_{UW}(u,w)dw = \int_{-\infty}^{\infty} f_{XY}(w, w-u)dw$$

Example 10: The joint p.d.f of two random variables X and Y is given by $f_{XY}(x,y)$. If we define the random variable $U = XY$, determine the p.d.f of U.

Solution: We define the auxiliary random variable $W = X$. Then the unique solution to the two equations is $x = w$ and $y = \frac{u}{x} = \frac{u}{w}$, and the Jacobian of the transformation is

$$J(x,y) = \begin{vmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial w}{\partial x} & \frac{\partial w}{\partial y} \end{vmatrix} = \begin{vmatrix} y & x \\ 1 & 0 \end{vmatrix} = -x = -w$$

Since there is only one solution to the equations, we have that

$$f_{UW}(u,w) = \frac{f_{XY}(x,y)}{|J(x,y)|} = \frac{1}{|w|} f_{XY}\left(w, \frac{u}{w}\right)$$

$$f_U(u) = \int_{-\infty}^{\infty} f_{UW}(u,w)dw = \int_{-\infty}^{\infty} \frac{1}{|w|} f_{XY}\left(w, \frac{u}{w}\right) dw$$

2.7

Correlation coefficient and Bivariate Normal Distribution

Meaning of correlation:

In a bivariate distribution we may be interested to find out if there is any **correlation** or **covariance** between the two variables under study. If the change in one variable affects a change in the other variable, the variables are said to be **correlated**. If the two variables deviate in the same direction, $i.e.$, if the increase (or decrease) in one results in a corresponding increase (or decrease) in the other, **correlation** is said to be **positive**. But, if they constantly deviate in the opposite directions, $i.e.$, if increase (or decrease) in one results in corresponding decrease (or increase) in the other, ***correlation*** *is said to be* ***negative***. For example, the correlation between (i) the heights and weights of a group of persons, and (ii) the income and expenditure; is positive and the correlation between (i) price and demand of a commodity and (ii) the volume and pressure of a perfect gas; is negative. *Correlation is said to be* ***perfect*** if the deviation in one variable is followed by a corresponding and proportional deviation in the other.

Karl Pearson's Coefficient of Correlation:

As a measure of intensity or degree of linear relationship between two variables, **Karl Pearson**, a British Biometrician developed a formula called **correlation coefficient**. Correlation coefficient between two variables X and Y, usually denoted by $\boldsymbol{\rho}(\boldsymbol{X}, \boldsymbol{Y})$ or $\boldsymbol{\rho}_{\boldsymbol{XY}}$, is a numerical measure of linear relationship between them and is defined by

$$\boldsymbol{\rho}(\boldsymbol{X}, \boldsymbol{Y}) = \frac{\boldsymbol{\sigma_{XY}}}{\boldsymbol{\sigma_X . \sigma_Y}} = \frac{\boldsymbol{cov\ (X,Y)}}{\sqrt{\boldsymbol{V(X)}}\,\sqrt{\boldsymbol{V(Y)}}}$$

where $\sigma_{XY} = cov\ (X, Y) = E\big[\big(X - E(X)\big)\ \big(Y - E(y)\big)\big]$,

$\sigma_X{}^2 = V(X) = E\big[\big(X - E(X)\big)\big]^2$ and $\sigma_Y{}^2 = V(Y) = E\big[\big(Y - E(Y)\big)\big]^2$

Note:

1. $\rho(X, Y)$ provides a measure of linear relationship between X and Y. For non – linear relationship, however, it is not suitable.

2. Karl Pearson's correlation coefficient is also called **product – moment correlation coefficient.**

Properties:

1. $-1 \leq \rho(X, Y) \leq 1$. If $\rho = -1$, the **correlation is perfect and negative**. If $\rho = 1$, the **correlation is perfect and positive**.
2. Correlation coefficient is independent of change of origin and scale. That is, if $U = \frac{X-a}{h}$ and $V = \frac{Y-b}{k}$, then $\rho(U, V) = \rho(X, Y)$

Theorem: Two independent variables are uncorrelated.

Proof:

Consider $\sigma_{XY} = cov\ (X, Y) = E\big[\big(X - E(X)\big)\big(Y - E(X)\big)\big]$

$\Longrightarrow\ \sigma_{XY} = E(X, Y) - E(X).E(Y)$ ...... (1)

If X and Y are independent, then

$E(XY) = E(X).E(Y)$ .... (2)

From (1) and (2), if X and Y are independent, then $\rho(X, Y) = 0$

The converse need not be true. That is, uncorrelated variables need not be independent.

Example 1 : Let $\boldsymbol{X \sim N(0, 1)}$ and $\boldsymbol{Y = X^2}$. Then $\boldsymbol{E(X) = E(X^3) = 0}$.

Solution: Consider $cov\ (X, Y) = E(XY) - E(X).E(Y) = E(X^3) - E(X).E(X^2)$

$$= 0 - 0 = 0$$

$\Longrightarrow cov\ (X, Y) = 0$ but X and Y are related by $Y = X^2$.

Thus, uncorrelated variables need not be independent.

Note: The converse is true if the joint distribution of (X, Y) is bivariate normal.

Example 2: The j.p.m.f of (X, Y) is given below:

$Y \backslash X$	-1	1
0	$\frac{1}{8}$	$\frac{3}{8}$
1	$\frac{2}{8}$	$\frac{2}{8}$

Find the correlation coefficient between X and Y

Solution : Computation of marginal p.m.fs

$Y \backslash X$	-1	1	$g(y)$
0	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{4}{8}$
1	$\frac{2}{8}$	$\frac{2}{8}$	$\frac{4}{8}$
$p(x)$	$\frac{3}{8}$	$\frac{5}{8}$	1

We have

$E(X) = \sum x\, p(x) = (-1) \times \frac{3}{8} + 1 \times \frac{5}{8} = -\frac{3}{8} + \frac{5}{8} = \frac{2}{8} = \frac{1}{4}$,

$E(X^2) = \sum x^2\, P(x) = (-1)^2 \frac{3}{8} + 1^2 \times \frac{5}{8} = \frac{3}{8} + \frac{5}{8} = 1$, then

$V(X) = E(X^2) - \left(E(X)\right)^2 = 1 - \left(\frac{1}{4}\right)^2 = 1 - \frac{1}{16} = \frac{15}{16}$

Similarly, $E(Y) = \sum y\, g(y) = 0 \times \frac{4}{8} + 1 \times \frac{4}{8} = \frac{4}{8} = \frac{1}{2}$

$E(Y^2) = \sum y^2\, g(y) = 0^2 \times \frac{4}{8} + 1^2 \times \frac{4}{8} = \frac{4}{8} = \frac{1}{2}$ and

$V(Y) = E(Y^2) - \left(E(Y)\right)^2 = \frac{1}{2} - \left(\frac{1}{2}\right)^2 = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$

Further, $E(XY) = 0 \times (-1) \times \frac{1}{8} + 0 \times 1 \times \frac{3}{8} + 1 \times (-1) \times \frac{2}{8} + 1 \times 1 \times \frac{2}{8} = 0$

Thus, $cov\,(X, Y) = E(XY) - E(X)E(Y)$

$$0 - \frac{1}{4} \times \frac{1}{2} = -\frac{1}{8}$$

$\therefore \rho\,(X, Y) = \frac{cov\,(X,Y)}{\sqrt{V(X)}\,\sqrt{V(Y)}} = -\frac{\frac{1}{8}}{\sqrt{\frac{15}{16} \times \frac{1}{4}}} = -\frac{1}{\sqrt{15}} = -0.2582$

Example 3: Two random variables X and Y have the joint probability density function

$$\boldsymbol{f(x,y)} = \begin{cases} \boldsymbol{2 - x - y} & , \quad \boldsymbol{0} < x < 1\,, \boldsymbol{0} < y < 1 \\ \boldsymbol{0} & , \quad \boldsymbol{otherwise} \end{cases}$$

Find correlation coefficient between X and Y.

Solution: By symmetry in x and y we have $f_1(x) = f_2(y)$, $E(X) = E(Y)$ and $V(X) = V(Y)$

The m.p.d.f X is given by

$$f_1(x) = \int_0^1 f(x, y)\, dy = \int_0^1 (2 - x - y)\, dy = \frac{3}{2} - x$$

Thus, $f_1(x) = \begin{cases} \frac{3}{2} - x & , \quad if \;\; 0 < x < 1 \\ \quad 0 & , \quad otherwise \end{cases}$

Consider.

$$E(X) = \int_0^1 x f_1(x) dx = \int_0^1 x\left(\frac{3}{2} - x\right) dx = \int_0^1 \left(\frac{3}{2}x - x^2\right) dx = \frac{5}{12}$$

$$E(X^2) = \int_0^1 x f_1(x) = \int_0^1 x^2\left(\frac{3}{2} - x\right) dx = \int_0^1 \left(\frac{3}{2}x^2 — x^3\right) dx = \frac{1}{4}$$

Further,

$$E(XY) = \int_0^1 \int_0^1 xy\, f(x,y) dx\, dy \ = \int_0^1 \int_0^1 xy\,(2 - x - y) dx\, dy$$

$$= \int_0^1 y\left(\int_0^1 (2x - x^2 - xy) dx\ \right) dy = \int_0^1 y\left[2.\frac{x^2}{2} - \frac{x^3}{3} - \frac{yx^2}{2}\right]_0^1 dy$$

$$= \int_0^1 y\left(1 - \frac{1}{3} - \frac{y}{2}\right) dy = \int_0^1 y\left(\frac{2}{3} - \frac{y}{2}\right) dy$$

$$= \int_0^1 \left(\frac{2}{3}y - \frac{y^2}{2}\right) dy = \left[\frac{y^3}{3} - \frac{y^3}{6}\right]_0^1 = \frac{1}{3} - \frac{1}{6} = \frac{1}{6}$$

$\therefore E(XY) = \dfrac{1}{6}$

Thus, $V(X) = E(X^2) - \left(E(X)\right)^2 = \frac{1}{4} - \left(\frac{5}{12}\right)^2 = \frac{1}{4} - \frac{25}{144} = \frac{36-25}{144} = \frac{11}{144}$

and $cov\ (X,Y) = E(XY) - E(X)E(Y) = \frac{1}{6} - \left(\frac{5}{12}\right)^2 = \frac{1}{6} - \frac{25}{144} = \frac{24-25}{144} = -\frac{1}{144}$

$\therefore$ The correlation coefficient is given by

$$\rho(X,Y) = \frac{cov\ (X,Y)}{\sqrt{V(x)}\ \sqrt{V(Y)}} = -\frac{\frac{1}{144}}{\sqrt{\frac{11}{144}}\sqrt{\frac{11}{144}}} = -\frac{\frac{1}{144}}{\frac{11}{144}} = -\frac{1}{11}$$

Bivariate Normal Distribution:

The bivaraite normal distribution is a generalization of a normal distribution for a single value.

Let X and Y be two normally correlated variables with correlation coefficient ρ. Let $E(X) = \mu_1,\ V(X) = \sigma_1^2, E(Y) = \mu_2$ and $V(Y) = \sigma_2^2$.

Definition: The bivariate continuous random variable (X, Y) is said to follow bivariate normal distribution with parameters $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$ and ρ if its p.d.f. is given by

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} exp\left[-\frac{1}{2(1-\rho^2)}\left\{\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right\}\right];$$

$$-\infty < x, y, \mu_1, \mu_2 < \infty,\ \ \sigma_1 > 0, \sigma_2 > 0 \text{ and } -1 < \rho < 1.$$

Notation: $(X, Y) \sim BN(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$. Read as (X, Y) follows **bivariate normal distribution** with parameters $\mu_1, \mu_2\sigma_1^2, \sigma_2^2$ and ρ.

Note: The curve $z = f(x, y)$ which is the equation of a surface in three dimensions is called the **Normal correlation surface**.

Marginal p.d.fs of X and Y: The m.p.d.f of X is given by

$$f_1(x) = \int_{-\infty}^{\infty} f(x,y)dy$$

Let $v = \frac{y-\mu_2}{\sigma_2}$, then $y = \mu_2 + \sigma_2 v$ and $dy = \sigma_2 dv$

Therefore,

$$f_1(x) = \frac{\sigma_2}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} exp\left[-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\, v\left(\frac{x-\mu_1}{\sigma_1}\right) + v^2\right\}\right] dv$$

$$= \frac{1}{2\pi\sigma_1\sqrt{1-\rho^2}} exp\left[-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right] \int_{-\infty}^{\infty} exp\left[-\frac{1}{2(1-\rho^2)}\left\{v - \rho\left(\frac{x-\mu_1}{\sigma_1}\right)\right\}^2\right] dv$$

Let $\frac{1}{\sqrt{1-\rho^2}}\left[v - \rho\left(\frac{x-\mu_1}{\sigma_1}\right)\right] = t$. Then $dv = \sqrt{1-\rho^2}\ dt$

$$\therefore\ f_1(x) = \frac{1}{2\pi\sigma_1} exp\left[-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)\right] \int_{-\infty}^{\infty} exp\left(-\frac{t^2}{2}\right) dt$$

$$= \frac{1}{2\pi\sigma_1} exp\left[-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right].\sqrt{2\pi}$$

$$\Rightarrow\ f_1(x) = \frac{1}{\sigma_1\sqrt{2\pi}} exp\left[-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right] \text{ for } -\infty < x < \infty$$

Similarly, it can be shown that

$$f_2(y) = \frac{1}{\sigma_2\sqrt{2\pi}} exp\left[-\frac{1}{2}\left(\frac{y-\mu_1}{\sigma_2}\right)^2\right] \text{ for } -\infty < x < \infty$$

Hence $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$.

Note: If $(X, Y) \sim BN(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho\ )$, then $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$

Conditional p.d.fs of X and Y

The conditional probability density function (c.p.d.f.) of X for given Y is given by

$$f_{1|2}(x|y) = \frac{f(x,y)}{f_2(y)}$$

$$= \frac{1}{\sigma_1\sqrt{2\pi}\sqrt{1-\rho^2}} exp\left[\frac{-1}{2(1-\rho^2)}\left\{\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\left(1-(1-\rho^2)\right)\right\}\right]$$

$$= \frac{1}{\sigma_1\sqrt{2\pi}\sqrt{1-\rho^2}}\ exp\left[\frac{-1}{2(1-\rho^2)\sigma_1^2}\left\{(x-\mu_1)^2 - 2\rho\frac{\sigma_1}{\sigma_2}(x-\mu_1)(y-\mu_2) + \frac{\sigma_1^2}{\sigma_2^2}\rho^2(y-\mu_2)^2\right\}\right]$$

$$= \frac{1}{\sigma_1\sqrt{2\pi}\sqrt{1-\rho^2}} exp\left[\frac{-1}{2(1-\rho^2)\sigma_1^2}\left\{(x-\mu_1) - \rho\frac{\sigma_1}{\sigma_2}(y-\mu_2)\right\}^2\right]$$

Therefore, $f_{1|2}(x|y) = \frac{1}{\sigma_1\sqrt{2\pi}\sqrt{1-\rho^2}} exp\left[\frac{-1}{2(1-\rho^2)\sigma_1^2}\left\{(x-\mu_1) - \rho\frac{\sigma_1}{\sigma_2}(y-\mu_2)\right\}^2\right]$

which is the univariate normal distribution with mean

$$E(X|Y=y) = \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(y-\mu_2) \text{ and}$$

$$V(X|Y=y) = \sigma_1^2(1-\rho^2)$$

Thus, the c.p.d.f of X for fixed Y is given by

$$(X|Y=y) \sim N\left[\mu_1+\rho\frac{\sigma_1}{\sigma_2}(y-\mu_2), \sigma_1^2(1-\rho^2)\right]$$

Similarly, the c.p.d.f of Y for fixed $X=x$ is given by

$$f_{2|1}(y|x)=\frac{1}{\sigma_2\sqrt{2\pi}\sqrt{1-\rho^2}}exp\left[\frac{-1}{2(1-\rho^2)\sigma_2^2}\left\{(y-\mu_2)-\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1)\right\}^2\right], -\infty<y<\infty$$

Thus, the c.p.d.f of Y for fixed X is given by

$$(Y|X=x) \sim N\left[\mu_2+\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1), \sigma_2^2(1-\rho^2)\right]$$

Example 4: If $(X,Y) \sim BN\ (5,10,1,25,\rho)$ where $\rho>0$, find ρ when $P(4<Y<16|X=5)=0.954$

Solution:

Here $\mu_1=5, \mu_2=10$, $\sigma_1^2=1, \sigma_2^2=25$. We know that $(Y|X=x) \sim N[\mu,\sigma^2]$

where $\mu=\mu_2+\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1)$ and $\sigma^2=\sigma_2^2(1-\rho^2)$.

Here $\mu=10+\rho\times\frac{5}{1}(5-5)=10$ and $\sigma^2=25(1-\rho^2)$

Thus $(Y|X=5) \sim N[10, 25(1-\rho^2)]$. We want to find ρ so that
$P(4<Y<16|X=5)=0.954$

Let $Z=\frac{Y-\mu}{\sigma}=\frac{Y-10}{5\sqrt{1-\rho^2}} \sim N(0,1) \Rightarrow P\left(\frac{4-10}{\sigma}<Z<\frac{16-10}{\sigma}\right)=0.954$

$$\Rightarrow P\left(-\frac{6}{\sigma}<Z<\frac{6}{\sigma}\right)=0.954 \Rightarrow P\left(0<Z<\frac{6}{\sigma}\right)=0.477$$

From standard normal table, we have $\frac{6}{\sigma}=2 \ \Rightarrow \sigma=3 \ \Rightarrow \sigma^2=9$

$$\Rightarrow 25(1-\rho^2)=9 \Rightarrow 1-\rho^2=\frac{9}{25} \Rightarrow \rho^2=1-\frac{9}{25}=\frac{16}{25} \Rightarrow \rho=\frac{4}{5}=0.8$$

Example 5: Find $cor(X,Y)$ for the jointly normal distribution

$$f(x,y) = \frac{1}{2\pi\sqrt{3}} exp\left[-\frac{\{(2x-y)^2+2xy\}}{6}\right], -\infty < x,y < \infty$$

Solution: Given $(X,Y) \sim BN(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$. Then its p.d.f. is given by

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} exp\left[-\frac{1}{2(1-\rho^2)}\left\{\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right\}\right] --(1)$$

We have

$$f(x,y) = \frac{1}{2\pi\sqrt{3}} exp\left[-\frac{\{(2x-y)^2+2xy\}}{6}\right], -\infty < x,y < \infty$$

$$i.e, \quad f(x,y) = \frac{1}{2\pi\sqrt{3}} exp\left[-\frac{\{4x^2+y^2-2xy\}}{6}\right] ----(2)$$

Comparing (1) and (2), we get $\mu_1 = \mu_2 = 0$. Then (1) becomes

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} exp\left[-\frac{\left\{\frac{x^2}{\sigma_1{}^2} + \frac{y^2}{\sigma_2{}^2} - 2\rho\frac{xy}{\sigma_1\sigma_2}\right\}}{2(1-\rho^2)}\right] — (3)$$

Comparing (2) and (3) , we find

$\sigma_1\sigma_2\sqrt{1-\rho^2} = \sqrt{3}$, $\sigma_1{}^2(1-\rho^2) = \frac{3}{4}$, $\sigma_2{}^2(1-\rho^2) = 3$ and $\frac{\rho}{\sigma_1\sigma_2(1-\rho^2)} = \frac{1}{3}$

On solving we get $\sigma_1{}^2 = 1$, $\sigma_2{}^2 = 4, \rho^2 = \frac{1}{4}$

Thus $cor(X,Y) = \rho = \sqrt{\frac{1}{4}} = \pm\frac{1}{2}$

Example 6: Determine the parameters of the bivariate normal distribution

$$f(x,y) = c\, exp\left[-\frac{\{16(x-2)^2 - 12(x-2)(y+3) + 9(y+3)^2\}}{216}\right]$$

Solution: If $(X,Y) \sim BN(\mu_1,\mu_2,\sigma_1{}^2,\sigma_2{}^2,\rho)$, then

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} exp\left[-\frac{\left\{\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right)^2 + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right\}}{2(1-\rho^2)}\right]$$

Comparing these functions, we get

$$\mu_1 = 2, \mu_2 = -3, c = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}, \frac{16}{216} = \frac{1}{2(1-\rho^2){\sigma_1}^2}$$

$$\frac{9}{216} = \frac{1}{2(1-\rho^2){\sigma_2}^2}, \frac{12}{216} = \frac{2\rho}{2\sigma_1\sigma_2(1-\rho^2)}$$

$\therefore (1-\rho^2){\sigma_1}^2 = \frac{27}{4}, (1-\rho^2){\sigma_2}^2 = 12, \sigma_1\sigma_2(1-\rho^2) = 18\rho$

$\Longrightarrow (1-\rho^2)^2{\sigma_1}^2{\sigma_2}^2 = 81 = (18\rho)^2 \Longrightarrow \rho^2 = \frac{1}{4},$

Further, $\sigma_1 = 3$ and $\sigma_2 = 4$.

Thus, $c = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} = \frac{1}{2\pi\times3\times4\sqrt{1-\frac{1}{4}}} = \frac{1}{12\pi\sqrt{3}}$

$\therefore (X,Y) \sim BN\left(2,3,9,16,\frac{1}{2}\right)$

Example 7: If $X \sim N(\mu,\sigma^2)$ and $(Y|x) \sim N(x,\sigma^2)$, show that

$$(X,Y) \sim BN(\mu,\mu,\sigma^2,2\sigma^2,\rho) .$$

Solution: We are given that

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right], -\infty < x < \infty$$

$$g(y|x) = \frac{1}{\sqrt{2\pi}\sigma} exp\left[-\frac{1}{2}\left(\frac{y-x}{\sigma}\right)^2\right], -\infty < y < \infty$$

$$\therefore h(x,y) = g(y|x)f(x) = \frac{1}{2\pi\sigma^2} exp\left[-\frac{1}{2}\left\{\left(\frac{x-\mu}{\sigma}\right)^2 + \left(\frac{y-x}{\sigma}\right)^2\right\}\right]$$

Consider $\left(\frac{y-x}{\sigma}\right)^2 = \left(\frac{y-\mu+\mu-x}{\sigma}\right)^2 = \left(\frac{y-\mu}{\sigma}\right)^2 + \left(\frac{x-\mu}{\sigma}\right)^2 - 2\left(\frac{x-\mu}{\sigma}\right)\left(\frac{y-\mu}{\sigma}\right)$

Thus, $h(x,y)=\frac{1}{2\pi\sigma^2}exp\left[-\frac{1}{2}\left\{2\left(\frac{x-\mu}{\sigma}\right)^2+\left(\frac{y-\mu}{\sigma}\right)^2-2\left(\frac{x-\mu}{\sigma}\right)\left(\frac{y-\mu}{\sigma}\right)\right\}\right]$

The bivariate normal p.d.f. is given by

$$f(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left[-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{x-\mu}{\sigma_1}\right)^2-2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right)+\left(\frac{y-\mu_2}{\sigma_2}\right)^2\right\}\right]$$

On comparing $h(x,y)$ with $f(x,y)$, we get

$$\sigma_1\sigma_2\sqrt{1-\rho^2}=\sigma^2 \quad , \quad {\sigma_1}^2(1-\rho^2)=\frac{1}{2}\sigma^2$$

$$\frac{\sigma_1\sigma_2(1-\rho^2)}{\rho}=\sigma^2 \quad , \quad {\sigma_2}^2(1-\rho^2)=\sigma^2\,, \quad \mu_1=\mu_2=\mu$$

On solving, we get $\rho^2=\frac{1}{2}\ \ {\sigma_2}^2=2\sigma^2\,,\ {\sigma_1}^2=\sigma^2$.

Thus, $(X,Y)\sim BN\left(\mu,\mu,\sigma^2,2\sigma^2,\frac{1}{\sqrt{2}}\right)$

Example 8: The variables X and Y are connected by the equation $aX+bY+c=0$. Show that the correlation between them is -1 if signs of a and b are same and $+1$ if they are different signs.

Solution: Given $aX+bY+c=0\Longrightarrow aE(X)+bE(Y)+c=0$

$\therefore a[X-E(X)]+b[Y-E(Y)]=0\Longrightarrow[X-E(X)]=-\frac{b}{a}[Y-E(Y)]$

$\therefore cov(X,Y)=E[\{X-E(X)\}\{Y-E(Y)\}]=-\frac{b}{a}E\big(Y-E(Y)\big)^2=-\frac{b}{a}{\sigma_Y}^2$ and

${\sigma_X}^2=E\big(X-E(X)\big)^2=\frac{b^2}{a^2}E\big(Y-E(Y)\big)^2=\frac{b^2}{a^2}{\sigma_Y}^2$

$$\therefore \rho=\frac{cov(X,Y)}{\sigma_X\cdot\sigma_Y}=\frac{-\frac{b}{a}{\sigma_Y}^2}{\sqrt{{\sigma_Y}^2}\sqrt{\frac{b^2}{a^2}{\sigma_Y}^2}}=\frac{-\frac{b}{a}{\sigma_Y}^2}{\left|\frac{b}{a}\right|{\sigma_Y}^2}=\frac{-\frac{b}{a}}{\left|\frac{b}{a}\right|}$$

$$\therefore \rho=\frac{cov(X,Y)}{\sigma_X\cdot\sigma_Y}=\begin{cases}1\ , & if\ a\ and\ b\ have\ opposite\ signs\\ -1\ , & if\ a\ and\ b\ have\ same\ signs\end{cases}$$

Unit – 3

Probability Inequalities and Generating Functions

3.1

Probability Inequalities

Inequalities are useful for bounding quanties that might otherwise be hard to compute. They will also be used in the **theory of convergence** and **limit theorems**.

Chebychev's Inequality

When we want to find the probability of an event described by a random variable, its c.d.f or p.d.f. or p.m.f. is required. If it is not known but its *mean* and *variance* are known, we can use **Chebychev's inequality** to find the **upper bound** or **lower bound** for the probability of the event.

Theorem 1: If X is a random variable with mean μ and variance σ^2, then

$$\boldsymbol{P(|X-\mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}} \qquad \dots\dots\dots\dots..(\mathbf{1})$$

where $\epsilon > 0$

Proof: The proof is given for a continuous random variable. Let X be a continuous r.v. with p.d.f. $f(x)$. Then

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

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

$$\geq \int_{-\infty}^{\mu-\epsilon}(x-\mu)^2 f(x)dx + \int_{\mu+\epsilon}^{\infty}(x-\mu)^2 f(x)dx$$

In the first integral,

$$x \leq \mu - \epsilon \Longrightarrow -x \geq -\mu + \epsilon \Longrightarrow -(x-\mu) \geq \epsilon \Longrightarrow (x-\mu)^2 \geq \epsilon^2$$

In the third integral, $x \geq \mu + \epsilon \Longrightarrow (x-\mu) \geq \varepsilon \Longrightarrow (x-\mu)^2 \geq \epsilon^2$

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

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

$$= \epsilon^2 P[\mu - \epsilon \geq X \geq \mu + \epsilon] = \epsilon^2 P[-\epsilon \geq X - \mu \geq \epsilon]$$

$$= \epsilon^2 P[|X-\mu| \geq \epsilon]$$

Thus, $\sigma^2 \geq \epsilon^2 P[|X-\mu| \geq \epsilon] \Longrightarrow P[|X-\mu| \geq \epsilon] \leq \frac{\sigma^2}{\epsilon^2}$

Note: The proof is similar as in the case of d.r.v.X except that integration is replaced by summation.

Alternative forms:

$$P[|X-\mu| < \epsilon] = 1 - P[|X-\mu| \geq \epsilon] \geq 1 - \frac{\sigma^2}{\epsilon^2}$$

$$\Longrightarrow \boldsymbol{P}[|\boldsymbol{X} - \boldsymbol{\mu}| < \epsilon] \geq \mathbf{1} - \frac{\boldsymbol{\sigma^2}}{\boldsymbol{\epsilon^2}} \quad \dots\dots\dots\dots.(2)$$

Let $\epsilon = k\sigma$ for $k > 0$. Then from (1), we have

$$\boldsymbol{P}[|\boldsymbol{X} - \boldsymbol{\mu}| \geq \boldsymbol{k\sigma}] \leq \frac{\mathbf{1}}{\boldsymbol{k^2}} \quad \dots\dots\dots\dots\dots..(3)$$

and from (2), we have

$$\boldsymbol{P}[|\boldsymbol{X} - \boldsymbol{\mu}| < k\sigma] \geq \mathbf{1} - \frac{\mathbf{1}}{\boldsymbol{k^2}} \quad \dots\dots\dots....(4)$$

Example 1: If a r.v. X has mean 12 and variance 9 and the probability distribution is unknown, then find $P(6 < X < 18)$.

Solution: Since the probability distribution of X is not known, we can't find the value of the required probability. We can find only a lower bound for probability using Chebychev's inequality. We have, for $\epsilon > 0$.

$$P[|X-\mu| < \epsilon] \geq 1 - \frac{\sigma^2}{\epsilon^2}$$

Given $E(X) = \mu = 12$ and $V(X) = \sigma^2 = 9$.

Then $P[|X-12| < \epsilon] \geq 1 - \frac{9}{\epsilon^2} \Rightarrow P[-\epsilon < (X-12) < \epsilon] \geq 1 - \frac{9}{\epsilon^2}$

$$\Rightarrow P[12-\epsilon < X < 12+\epsilon] \geq 1 - \frac{9}{\epsilon^2}$$

Let $\epsilon = 6$. Then $P[6 < X < 18] \geq 1 - \frac{9}{36} = 1 - \frac{1}{4} = 0.75$

$$\Rightarrow P[6 < X < 18] \geq 0.75$$

Thus, the probability of X lying between 6 and 18 is alseast 75%.

Example 2: A d.r.v. X takes the values $-1, 0$ and 1 with probabilities $\frac{1}{8}, \frac{3}{4}$ and $\frac{1}{8}$ respectively. Evaluate $P[|X-\mu| \geq 2\sigma]$ and compare it with the upper bound given by Chebychev's inequality.

Solution: We have,

X	-1	0	1
$p(x)$	$\frac{1}{8}$	$\frac{3}{4}$	$\frac{1}{8}$

Then $E(X) = \mu = \sum xp(x) = -1 \times \frac{1}{8} + 0 \times \frac{3}{4} + 1 \times \frac{1}{8} = 0$

and $E(X^2) = \sum x^2 p(x) = 1 \times \frac{1}{8} + 0 \times \frac{3}{4} + 1 \times \frac{1}{8} = \frac{2}{8} = \frac{1}{4}$

Hence $\sigma^2 = V(X) = E(X^2) - \left(E(X)\right)^2 = \frac{1}{4} - 0 = \frac{1}{4}$

Consider $P[|X-\mu| \geq 2\sigma] = P\left[|X| \geq 2 \cdot \frac{1}{2}\right] = P[|X| \geq 1]$

$$= P(X = -1,1)$$

$$= P(X = -1) + P(X = 1)$$

$$= \frac{1}{8} + \frac{1}{8} = \frac{2}{8} = \frac{1}{4} = 0.25$$

$\Longrightarrow P[|X-\mu| \geq 2\sigma] \leq 0.25$

On the other hand, by Chebychev's inequality,

$$P[|X-\mu| \geq 2\sigma] \leq \frac{1}{2^2} = \frac{1}{4}$$

Note that the two values are same.

Example 3: Use Chebychev's inequality to find how many times must a fair coin be tossed in order that the probability that the ratio of the number of heads to the number of tosses will lie between $\mathbf{0.45}$ and $\mathbf{0.55}$ will be at least $\mathbf{0.95}$.

Solution: Let X denote the number of heads obtained when a fair coin is tossed n times. Then $X \sim B(n,p)$. That is $E(X) = np$ and $V(X) = npq$.

Let $Y = \frac{X}{n}$. Then $E(Y) = E\left(\frac{X}{n}\right) = \frac{1}{n}E(X) = \frac{np}{n} = p$

and $V(Y) = V\left(\frac{X}{n}\right) = E\left(\left(\frac{X}{n}\right)^2\right) - \left(E\left(\frac{X}{n}\right)\right)^2 = \frac{1}{n^2}(E(X^2) - \left(E(X)\right)^2$

$$= \frac{1}{n^2}V(X) = \frac{npq}{n^2} = \frac{pq}{n}.$$

Since $p = \frac{1}{2}$ for a fair coin, $E(Y) = \frac{1}{2}$ and $V(Y) = \frac{1}{4n}$

By Chebychev's inequality for Y

$$P\left[\left|Y - \frac{1}{2}\right| < \epsilon\right] \geq 1 - \frac{\frac{1}{4n}}{\epsilon^2} = 1 - \frac{1}{4n\epsilon^2}$$

$$\Longrightarrow P\left[\frac{1}{2} - \epsilon < Y < \frac{1}{2} + \epsilon\right] \geq 1 - \frac{1}{4n\epsilon^2}$$

Notice that, if $\epsilon = 0.05$ then $P(0.45 < Y < 0.55) \geq 1 - \frac{1}{4n\epsilon^2}$

Now, find n when $\epsilon = 0.05$ and $1 - \frac{1}{4n\epsilon^2} = 0.95 \Longrightarrow 1 - \frac{1}{n\times 4\times(0.05)^2} = 0.95$

$\Longrightarrow 1 - \frac{1}{0.01\times n} = 0.95 \Longrightarrow \frac{1}{0.01\times n} = 0.05 \Longrightarrow n = \frac{1}{0.01\times 0.05} = \frac{1}{0.0005} = \frac{10000}{5} = 2000$

Thus, $n = 2000$

Bienayme – Chebychev's inequality

Theorem 3: Let $g(X)$ be a non-negative function of a r.v. X . Then for every $k > 0$, we have

$$\boldsymbol{P[g(X) \geq k] \leq \frac{E[g(X)]}{k}} \dots\dots\dots\dots.(\mathbf{1})$$

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

Let S be the set of all X for which $g(X) \geq k$.That is, $S = \{x \mid g(x) \geq k\}$.

Now, $E\left[g(X)\right] = \int_S g(x) f(x) dx,$ where $f(x)$ is the p.d.f. of X

$$\geq k\int_S f(x)dx \quad \left(on\ S, g(x) \geq k\right)$$

$$= kP\left[g(X) \geq k\right]$$

$$\Rightarrow P\left[g(X) \geq k\right] \leq \frac{E\left[g(X)\right]}{k}$$

Note:

1. If $g(X) = \left(X - E(X)\right)^2 = (X - \mu)^2$, then $E\left(g(X)\right) = V(X) = \sigma^2$ and replacing k by $\epsilon^2\sigma^2$ in equation (1), we get

$$\boldsymbol{P[(X-\mu)^2 \geq \epsilon^2\sigma^2] \leq \frac{\sigma^2}{\epsilon^2\sigma^2} = \frac{1}{\epsilon^2}}$$

$$\Rightarrow \boldsymbol{P[|X-\mu| \geq \epsilon\sigma] \leq \frac{1}{\epsilon^2}}$$

which is **Chebychev's inequality.**

2. If $g(X) = |X|$ in (1), then we get for any $k > 0$,

$$\boldsymbol{P[|X| \geq k] \leq \frac{E(|X|)}{k}}$$

which is known as **Markov's inequality.**

3. If $g(X) = |X|^r$ in (1), then we get

$$\boldsymbol{P[|X|^r \geq k^r] \leq \frac{E(|X|^r)}{k^r}}$$

which is known as **generalized Markov's inequality.**

Cauchy – Schwartz Inequality

When the j.p.d.f. of X and Y is known, upper bound for expected value of the product of X and Y can be found by using Cauchy – Schwartz inequality when second moments about origin of X and Y are given ($i.e.$, $E(X^2)$ and $E(Y^2)$ are given).

Theorem 2: For any two random variables $\boldsymbol{X}$ and $\boldsymbol{Y}$

$$\boldsymbol{\left(E(XY)\right)^2 \leq E(X^2)E(Y^2)}$$

Proof: Consider $E(X - tY)^2 \geq 0$ for any real number t. That is,

$E(X^2 - 2tXY + t^2Y^2) = E(X^2) - 2tE(XY) + t^2E(Y^2) \geq 0$
which is a quadratic expression in t. This expression is always positive only when t

has complex roots. This is possible only when discriminant of the expression is negative. Thus,

$$4\left(E(XY)\right)^2 - 4E(X^2)E(Y^2) \leq 0$$

$$\Rightarrow \left(E(XY)\right)^2 \leq E(X^2)E(Y^2)$$

Hence the result.

Example 4: The j.p.d.f. of (X, Y) is given by

$$f(x,y) = \frac{x+y}{21} \text{ for } x = 1,2,3 \text{ and } y = 1,2.$$

Verify whether Cauchy-Schwartz inequality.

Solution: The joint and marginal p.m.fs $f_1(x)$ and $f_2(y)$ of X and Y respectively are given in the following table.

$Y \backslash X$	1	2	3	$f_2(y)$
1	$\frac{2}{21}$	$\frac{3}{21}$	$\frac{4}{21}$	$\frac{9}{21}$
2	$\frac{3}{21}$	$\frac{4}{21}$	$\frac{5}{21}$	$\frac{12}{21}$
$f_1(x)$	$\frac{5}{21}$	$\frac{7}{21}$	$\frac{9}{21}$	1

$$E(X) = \sum_{x=1}^{3} x f_1(x) = 1 \times \frac{5}{12} + 2 \times \frac{7}{21} + 3 \times \frac{9}{21} = \frac{46}{21}$$

$$E(X^2) = \sum_{x=1}^{3} x^2 f(x) = 1^2 \times \frac{5}{12} + 2^2 \times \frac{7}{21} + 3^2 \times \frac{9}{21} = \frac{114}{21}$$

$$E(Y) = \sum_{y=1}^{2} y f_2(y) = 1 \times \frac{9}{21} + 2 \times \frac{12}{21} = \frac{33}{21}$$

$$E(Y^2) = \sum_{y=1}^{2} y^2 f_2(y) = 1^2 \times \frac{9}{21} + 2^2 \times \frac{12}{21} = \frac{57}{21}$$

$$E(XY) = \sum_{x=1}^{3} \sum_{y=1}^{2} xyf(x,y)$$

$$= 1 \times 1 \times \frac{2}{21} + 1 \times 2 \times \frac{3}{21} + 1 \times 3 \times \frac{4}{21} + 2 \times 1 \times \frac{3}{21} + 2 \times 2 \times \frac{4}{21} + 2 \times 3 \times \frac{5}{21}$$

$$= \frac{1}{21}(2 + 6 + 12 + 6 + 16 + 30) = \frac{71}{21}$$

Verification of Cauchy-Schwatz inequality:

Here $\left(E(XY)\right)^2 = \left(\frac{71}{21}\right)^2 = 11.755$ and $E(X^2)E(Y^2) = \frac{114}{21} \times \frac{57}{21} = 14.735$

Note that $\left(E(XY)\right)^2 \leq E(X^2)E(Y^2)$.

Example 5: Let X and Y be c.r.vs with j.p.d.f.

$$\boldsymbol{f(x,y) = \begin{cases} x + y, & 0 < x < 1, 0 < y < 1 \\ 0, & otherwise \end{cases}}$$

Verify Cauchy-Schwartz inequality.

Solution: The m.p.d.f. of X is given by

$$f_1(x) = \int_{-\infty}^{\infty} f(x,y)dy = \int_0^1 (x+y)dy = \left[xy + \frac{y^2}{2}\right]_0^1 = x + \frac{1}{2}$$

$$f_1(x) = \begin{cases} x + \frac{1}{2}, & 0 < x < 1 \\ 0, & otherwise \end{cases}$$

Since $f(x,y)$ is symmetric in $x \; and \; y$, the m.p.d.f. of Y is given by

$$f_2(y) = \begin{cases} y + \frac{1}{2}, & 0 < x < 1 \\ 0, & otherwise \end{cases}$$

Now,

$$E(X^2) = \int_0^1 x^2\left(x + \frac{1}{2}\right)dx = \int_0^1 \left(x^3 + \frac{x^2}{2}\right)dx = \left[\frac{x^4}{4} + \frac{x^3}{6}\right]_0^1 = \frac{1}{4} + \frac{1}{6} = \frac{10}{24}$$

Similarly, $E(Y^2) = \frac{10}{24}$ .Now,

$$E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)dxdy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy(x+y)dxdy$$

$$= \int_0^1 \left\{\int_0^1 (x^2y + xy^2)dx\right\}dy = \int_0^1 \left(\frac{x^3y}{3} + \frac{x^2y^2}{2}\right)_0^1 dy$$

$$= \frac{1}{6}\int_0^1 (2y + 3y^2)dy = \left[\frac{y^2}{6} + \frac{y^3}{6}\right]_0^1 = \frac{2}{6} = \frac{1}{3}$$

Thus $\left(E(XY)\right)^2 = \left(\frac{1}{3}\right)^2 = \frac{1}{9} = 0.111$, and $E(X^2)E(Y^2) = \frac{10}{24} \times \frac{10}{24} = 0.1736$

Hence $\left(E(XY)\right)^2 \leq E(X^2)E(Y^2)$.

3.2

Moment Generating Function

Certain derivations presented in modules 2.4, 2.5 and 2.6 have been somewhat heavy on algebra. For example, determining the mean and variance of the **Binomial distribution** turned out to be fairly tiresome. Another example of hard work was determining the set of probabilities associated with a sum , $P(X+Y=t)$. Many of these tasks are greatly simplified by using **probability generating functions.**

Moment Generating Function: The moment generating function (m.g.f) of a random variable X is dentoed by $M_X(t)$ and it is defined as

$$\boldsymbol{M_X(t) = E(e^{tX})}$$

$$\therefore\ M_X(t) = E(e^{tX}) = E\left[1 + \frac{tX}{1!} + \frac{(tX)^2}{2!} + \frac{(tX)^3}{3!} + \cdots\right]$$

$$= E(1) + \frac{t}{1!}E(X) + \frac{t^2}{2!}E(X^2) + \frac{t^3}{3!}E(X^3) + \cdots + \frac{t^r}{r!}E(X^r) + \cdots + \infty$$

$$\therefore\ M_X(t) = 1 + \frac{t}{1!}\mu_1' + \frac{t^2}{2!}\mu_2' + \frac{t^3}{3!}\mu_3' + \cdots + \frac{t^r}{r!}\mu_r' + \cdots + \infty \qquad \ldots\ldots \quad (1)$$

$$M_X(t) = \sum_{r=0}^{\infty} \frac{t^r}{r!}\mu_r'$$

which gives the m.g.f in terms of moments.

Therefore the coeffceint of $\frac{t^r}{r!}$ in $M_X(t)$ is μ_r' , where $r = 1, 2, 3, \ldots$ and $\mu_r' = E(X^r)$, moment about origin.

The m.g.f of X about mean $\mu = \mu_1' = E(X)$ is defined as

$$M_{X-\mu}(t) = E\left[e^{t(X-\mu)}\right] = E\left[1 + \frac{t}{1!}(X-\mu) + \frac{t^2}{2!}(X-\mu)^2 + \frac{t^3}{3!}(X-\mu)^3 + \cdots\right]$$

$$= 1 + \frac{t}{1!}E(X-\mu) + \frac{t^2}{2!}E(X-\mu)^2 + \frac{t^3}{3!}E(X-\mu)^3 + \cdots$$

$$= 1 + \frac{t}{1!}\mu_1 + \frac{t^2}{2!}\mu_2 + \frac{t^3}{3!}\mu_3 + \cdots$$

where $E(x-\mu)^r = \mu_r$ is known as the r^{th} central moment for $r = 1, 2, \ldots$

Note that $\mu_1 = E(X-\mu) = E(X) - \mu = \mu - \mu = 0$

Since $M_X(t)$ generates moments, it is called **moment generating function**.

If X is a discrete random variable with p.m.f. $p(x)$ then

$$M_X(t) = E\left(e^{tX}\right) = \sum_x e^{tx} p(x)$$

If X is a continuous random variable with p.d.f. $f(x)$, then

$$M_X(t) = E\left(e^{tX}\right) = \int_{-\infty}^{\infty} e^{tx} f(x)\,dx$$

Moments Using Moment Generating Function:

Differentiating equation (1) with respect to t and then putting $t = 0$, gives

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

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

In general,

$$\mu_r' = \left[\frac{d^r}{dt^r} M_X(t)\right]_{t=0}, \quad r = 1, 2, 3, \ldots$$

Note: Moment generating function $M_X(t)$ is used to calculate the higher moments.

Theorems on Moment Generating Function:

Theorem 1: $\boldsymbol{M_{aX}(t) = M_X(at)}$**, where** $\boldsymbol{a}$ **is a constant.**

Proof: By definition $M_{aX}(t) = E(e^{taX}) = E(e^{atX}) = M_X(at)$

Therefore, $M_{aX}(t) = M_X(at)$

Theorem 2: The moment generating function of the sum of n independent random variables is equal to the product of their respective moment generating functions, $\boldsymbol{i.e.}$, $\boldsymbol{M_{X_1+X_2+X_3+\cdots+X_n}(t) = M_{X_1}(t)M_{X_2}(t)M_{X_3}(t)\ldots M_{X_n}(t)}$

Proof: By definition,

$$M_{X_1+X_2+X_3+\cdots+X_n}(t) = \left[E^{t(X_1+X_2+X_3+\cdots+X_n)}\right]$$

$$= E(e^{tX_1})E(e^{tX_2})E(e^{tX_3})\ldots E(e^{tX_n})$$

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

Therefore, $M_{X_1+X_2+X_3+\cdots+X_n}(t) = M_{X_1}(t)M_{X_2}(r)\ldots M_{X_n}(t)$
Hence the proof.

Uniqueness Theorem of Moment Generating Function:

The m.g.f. of a distribution, if exists, uniquely determines the distribution. This implies that corresponding to a given probability distribution, there is only one m.g.f (provided it exists) and corresponding to a given m.g.f, there is only one probability distribution. Hence $M_X(t) = M_Y(t) \Longrightarrow X$ and Y are identically distributed.

Effect of Change of Origin and Scale on Moment Generating Function:

Let a random variable X be transformed to a new variable U by changing both the origin and scale in X as $= \frac{X-a}{h}$, where a and h are constants.

The m.g.f of U (about origin) is given by

$$M_U(t) = E(e^{tU}) = E\left[e^{t\left(\frac{X-a}{h}\right)}\right] = E\left(e^{\left(\frac{tX}{h}-\frac{ta}{h}\right)}\right) = e^{-\frac{at}{h}}E\left(e^{\left(\frac{tX}{h}\right)}\right)$$

$$\therefore\ \boldsymbol{M_{\frac{X-a}{h}}(t) = e^{-\frac{at}{h}}M_X\left(\frac{t}{h}\right)}$$

Note: If $Y = aX + b$, then $M_Y(t) = e^{bt}M_X(at)$

Example 1: If X represents the outcome when a fair die is tossed, find the m.g.f. of X and hence, find $E(X)$ and $Var(X)$.

Solution: When a fair die is tossed

$$P(X = x) = \frac{1}{6}, \quad x = 1, 2, 3, 4, 5, 6$$

$$\therefore \; M_X(t) = \sum_{X=1}^{6} e^{tx} P(X = x) = \frac{1}{6} \sum_{x=1}^{6} e^{tx}$$

$$= \frac{1}{6}(e^t + e^{2t} + e^{3t} + e^{4t} + e^{5t} + e^{6t})$$

$$E(X) = \left[\frac{d}{dt} M_X(t)\right]_{t=0} = \frac{1}{6}[e^t + 2e^{2t} + 3e^{3t} + 4e^{4t} + 5e^{5t} + 6e^{6t}]_{t=0}$$

$$= \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = \frac{21}{6} = \frac{7}{2}$$

$\therefore$ Mean $= E(X) = \frac{7}{2}$

Now, $E[X^2]_{t=0} = \left\{\frac{d^2}{dt^2}[M_X(t)]\right\}_{t=0}$

$$= \frac{1}{6}[e^t + 4e^{2t} + 9e^{3t} + 16e^{4t} + 25e^{5t} + 36e^{6t}]_{t=0}$$

$$= \frac{1}{6}(1 + 4 + 9 + 16 + 25 + 36) = \frac{91}{6}$$

$V(X) = E(X^2) - \left(E(X)\right)^2 = \frac{91}{6} - \frac{49}{4} = \frac{35}{12}.$

Example 2: Find the m.g.f. of the random variable X whose probability function $\boldsymbol{P(X = x) = \frac{1}{2^x}, x = 1, 2, 3, \ldots}$ and hence find its mean.

Solution: By definition,

$$M_X(t) = E\left(e^{tX}\right) = \sum_{x=0}^{\infty} e^{tx} P(X = x) = \sum_{x=1}^{\infty} e^{tX}\left(\frac{1}{2^x}\right) = \sum_{x=1}^{\infty}\left(\frac{e^t}{2}\right)^x$$

$$= \left[\frac{e^t}{2} + \left(\frac{e^t}{2}\right)^2 + \left(\frac{e^t}{2}\right)^3 + \cdots\right]$$

$$= \frac{e^t}{2}\left[1 + \frac{e^t}{2} + \left(\frac{e^t}{2}\right)^2 + \cdots\right] = \frac{e^t}{2}\left(1 - \frac{e^t}{2}\right)^{-1}$$

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

Therefore, $M_X(t) = \frac{e^t}{2-e^t}$

$$\mu_1' = \left[\frac{d}{dt}M_X(t)\right]_{t=0} = \left[\frac{d}{dt}\left(\frac{e^t}{2-e^t}\right)\right]_{t=0} = \left[\frac{(2-e^t)e^t - e^t(-e^t)}{(2-e^t)^2}\right]_{t=0} = \frac{(2-1)1+1}{(2-1)^2} = 2$$

Thus, $E(X)$ =mean= 2

Example 3: If the moments of a random variable X are defined by $E(X^r) = 0.6, r = 1, 2, \ldots$ .Show that $P(X = 0) = 0.4$, $P(X = 1) = 0.6$, and $P(X \geq 2) = 0$.

Solution: We know that

$$M_X(t) = \sum_{r=0}^{\infty} \frac{t^r}{r!}\mu_r'$$

where $\mu_r' = E(X^r) = 0.6$

$$\therefore\ M_X(t) = 1 + \sum_{r=1}^{\infty}\frac{t^r}{r!}\mu_r' = 1 + \sum_{r=1}^{\infty}\frac{t^r}{r!}(0.6) = 1 + (0.6)\left(\frac{t}{1!} + \frac{t^2}{2!} + \frac{t^3}{3!} + \cdots\right)$$

$$= 1 + (0.6)(e^t - 1) = 1 - 0.6 + 0.6e^t = 0.4 + 0.6e^t \qquad \ldots\ldots(1)$$

But by definition,

$$M_X(t) = E\left(e^{tx}\right) = \sum_{r=0}^{\infty} e^{tx}P(X = x)$$

$M_X(t) = P(X = 0) + e^tP(X = 1) + e^{2t}P(X = 2) + e^{3t}P(X = 3) +$...+... ...... (2)

From equations (1) and (2), we have

$$0.4 + 0.6e^t = P(X = 0) + e^t P(X = 1) + e^{2t} P(X = 2) + e^{3t} P(X = 3) + \cdots$$

Equating the coefficients of like terms on both sides,

$$P(X = 0) = 0.4,\ P(X = 1) = 0.6$$

$P(X = 2) = P(X = 3) = \cdots = 0 \quad \Longrightarrow P(X \geq 2) = 0$

Example 4: Find the m.g.f. of a random variable whose moments are $\boldsymbol{\mu_r = (r+1)!\ 2^r}$.

Solution: By definition, we have $M_X(t) = \sum_{r=0}^{\infty} \frac{t^r}{r!}\, \mu_r'$

$$= \sum_{r=0}^{\infty} \frac{t^r}{r!}\,(r+1)!\,2^r = \sum_{r=0}^{\infty} (r+1)(2t)^r$$

$$= 1 + 2(2t) + 3(2t)^2 + \cdots = (1-2t)^{-2} = \frac{1}{(1-2t)^2}$$

$\therefore \quad M_X(t) = \frac{1}{(1-2t)^2}$

Example 5: If $\boldsymbol{X \sim B(n, p)}$, find the m.g.f of $\boldsymbol{X}$ and hence find its mean and variance.

Solution: Since $X \sim B(n,p)$, its p.m.f. is given by

$p(x) = \binom{n}{x} p^x q^{n-x},\ x = 0, 1, 2, \ldots, n$ and $q = 1 - p$.

Then the m.g.f. of X is given by

$$M_X(t) = E\left[e^{tX}\right] = \sum_{x=0}^{n} e^{tx} p(x) = \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x q^{n-x} = \sum_{x=0}^{n} \binom{n}{x} \left(pe^t\right)^x q^{n-x}$$

$\Longrightarrow M_X(t) = (q + pe^t)^n$

$\frac{d}{dt} M_X(t) = n(q + pe^t)^{n-1} pe^t \Longrightarrow$Mean$= \mu_1' = \left[\frac{dM_X(t)}{dt}\right]_{t=0} = np$

Next, $\frac{d^2}{dt^2}\left(M_X(t)\right) = np[(n-1)(q+pe^t)^{n-2}pe^{2t} + (q+pe^t)^{n-1}e^t]$

$\Longrightarrow \ \mu_2' = \left[\frac{d^2}{dt^2}\left(M_X(t)\right)\right]_{t=0} = np[(n-1)p+1] = np[np-p+1] = np(np+q)$

$\Longrightarrow \ \mu_2' = n^2p^2 + npq$

Now, variance $= \sigma^2 = \mu_2' - (\mu_1')^2 = n^2p^2 + npq - n^2p^2 = npq$

Thus, $\mu = np$ and $\sigma^2 = npq$

Note that $\sigma^2 = npq = \mu q$ where $(0 < q < 1)$ .Thus, $\mu > \sigma^2$.

Note: *For binomial distribution, mean is always greater than variance.*

Example 6 : If $X \sim P(\lambda)$, find its m.g.f. and hence find its mean and variance.

Solution: Since $X \sim P(\lambda)$, then its p.m.f. is given by

$$p(x) = P(X = x) = \frac{e^{-\lambda}\lambda^x}{\lambda!} \ , x = 0, 1, \dots \text{ and } \lambda > 0$$

The m.g.f. of X is given by

$$M_X(t) = E\left[e^{tX}\right] = \sum_{x=0}^{\infty} e^{tx} p(x) = \sum_{x=0}^{\infty} e^{tx} \frac{e^{-\lambda}\lambda^x}{x!}$$

$$= e^{-\lambda} \sum_{x=0}^{\infty} \frac{\left(\lambda e^t\right)^x}{x!} = e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t - 1)}$$

$\Longrightarrow \ M_X(t) = e^{\lambda(e^t-1)}$

Since $\frac{d}{dt}\left(M_X(t)\right) = e^{\lambda(e^t-1)}\lambda e^t$; Mean$= \mu = \mu' = \left[\frac{d}{dt}\left(M_X(t)\right)\right]_{t=0} = \lambda.$

Now, $\frac{d^2}{dt^2}\left(M_X(t)\right) = \lambda\left[e^{\lambda(e^t-1)}e^t + e^{\lambda(e^t-1)}\lambda e^{2t}\right]$

Then $\mu_2' = \left[\frac{d^2 M_X(t)}{dt^2}\right]_{t=0} = \lambda(1+\lambda) = \lambda + \lambda^2$

Thus, variance $= \sigma^2 = \mu_2' - (\mu_1')^2 = \lambda + \lambda^2 - \lambda^2 = \lambda$

Therefore, $\mu = \sigma^2 = \lambda$

Note: *Mean and variance are same for Poisson distribution.*

Example 7: If $X \sim NB(r, p)$, find its m.g.f. and hence find its mean and variance.

Solution*:* Since $X \sim NB(r, p)$, its p.m.f. is given by

$$p(x) = \binom{-r}{x} p^r (-q)^x, \qquad x = 0, 1, 2, \dots$$

The m.g.f of X is given by

$$M_X(t) = \sum_{x=0}^{\infty} e^{tx} \binom{-r}{x} p^r (-q)^x$$

$$= \sum_{x=0}^{\infty} \binom{-r}{x} p^r \left(-qe^t\right)^x = p^r \sum_{x=0}^{\infty} \binom{-r}{x} \left(-qe^t\right)^x$$

$\Rightarrow \quad M_X(t) = p^r (1 - qe^t)^{-r}.$

Now, $\frac{d}{dt}\big(M_X(t)\big) = p^r(-r)(1 - qe^t)^{-(r+1)}(-qe^t) = qrp^r(1 - qe^t)^{-(r+1)}e^t$

Mean= $\mu = \mu_1' = \left[\frac{d}{dt}\big(M_X(t)\big)\right]_{t=0} = qrp^r(1-q)^{-(r+1)} = qrp^r(p)^{-(r+1)} = \frac{rq}{p}$

Further, $\frac{d^2}{dt^2}\big(M_X(t)\big) = rqp^r \frac{d}{dt}\left\{(1 - qe^t)^{-(r+1)}e^t\right\}$

$$= rqp^r\left\{-(r+1)(1 - qe^t)^{-(r+2)}(-qe^{2t}) + (1 - qe^t)^{-(r+1)}e^t\right\}$$

Then $\mu_2' = \left[\frac{d^2}{dt^2}\big(M_X(t)\big)\right]_{t=0} = rqp^r\left[(r+1)qp^{-(r+2)} + p^{-(r+1)}\right]$

$$= r(r+1)q^2p^{-2} + rqp^{-1} = \frac{rq}{p}\left(\frac{(r+1)q}{p} + 1\right) = \frac{rq}{p^2}(rq + 1)$$

$\Rightarrow \quad \mu_2' = \dfrac{r^2q^2}{p^2} + \dfrac{rq}{p^2}$

Hence, variance $= \sigma^2 = \mu_2' - (\mu_1')^2 = \frac{r^2q^2}{p^2} + \frac{rq}{p^2} - \frac{r^2q^2}{p^2} \Rightarrow \quad \sigma_2^2 = \frac{rq}{p^2}$

Example 8:Let X be a random variable with p.d.f.

$$f(x) = \begin{cases} \frac{1}{3} e^{-\frac{x}{3}} & , \quad x > 0 \\ 0 & , \quad otherwise \end{cases}$$

Find

(i) $P(X > 3)$

(ii) **M.g.f. of X**

(iii) $E(X)$ **and** $Var(X)$

Solution:

(i) $$P(X > 3) = \int_3^\infty f(x)\, dx = \int_3^\infty \frac{1}{3} e^{-\frac{x}{3}} dx = \frac{1}{3}\left[\frac{e^{-\frac{x}{3}}}{-\frac{1}{3}}\right]_3^\infty = -(0 - e^{-1}) = e^{-1} = \frac{1}{e}$$

(ii) $$M_X(t) = E\left(e^{tX}\right) = \int_0^\infty e^{tx} f(x)\, dx$$

$$= \int_0^\infty e^{tx} \frac{1}{3} e^{-\frac{x}{3}} dx = \frac{1}{3}\int_0^\infty e^{\left(t-\frac{1}{3}\right)x} dx = \frac{1}{3}\int_0^\infty e^{-\left(\frac{1}{3}-t\right)x} dx = \frac{1}{3}\left[\frac{e^{-\left(\frac{1}{3}-t\right)x}}{-\left(\frac{1}{3}-t\right)}\right]_0^\infty$$

$$= \frac{1}{3}\left[0 - \frac{1}{-\left(\frac{1}{3}-t\right)}\right] = \frac{1}{3}\left[\frac{1}{\left(\frac{1-3t}{3}\right)}\right]$$

$$M_X(t) = \frac{1}{1-3t} = (1-3t)^{-1}$$

$$\frac{d}{dt}[M_X(t)] = -(1-3t)^{-2}(-3) = 3(1-3t)^{-2}$$

(iii) $$E(X) = Mean = \left[\frac{d}{dt} M_X(t)\right]_{t=0} = 3$$

$$\frac{d^2}{dt^2}[M_X(t)] = -6(1-3t)^{-3}(-3) = 18(1-3t)^{-3}$$

$$E(X^2) = \left[\frac{d^2}{dt^2} M_X(t)\right]_{t=0} = 18$$

$$Var(X) = E(X^2) - [E(X)]^2 = 18 - 9 = 9$$

Example 9: Let X be a discrete random variable with p.d.f.

$$\boldsymbol{p(x) = \begin{cases} \dfrac{1}{x(x+1)} & , \quad x = 1, 2 \ldots \\ 0 & , \quad otherwise \end{cases}}$$

Show that $\boldsymbol{E(X)}$ does not exist even though m.g.f. exist.

Solution:

$$E(X) = \sum_{x=1}^{\infty} x\, p(x) = \sum_{x=1}^{\infty} \frac{1}{x+1} = \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \ldots = \sum_{x=1}^{\infty} \frac{1}{x} - 1$$

But $\boldsymbol{\sum_{x=1}^{\infty} \frac{1}{x}}$ is a divergent series.

Therefore, $E(x)$ does not exist and hence, no moment exists.

Now, m.g.f. of X is given by

$$M_X(t) = \sum_{x=1}^{\infty} p(x) e^{tx} = \sum_{x=1}^{\infty} \frac{e^{tx}}{x(x+1)}$$

Substituting $z = e^t$,

$$M_X(t) = \sum_{x=1}^{\infty} \frac{z^x}{x(x+1)} = \frac{z}{1.2} + \frac{z^2}{2.3} + \frac{z^3}{3.4} + \ldots$$

$$= z\left(1 - \frac{1}{2}\right) + z^2\left(\frac{1}{2} - \frac{1}{3}\right) + z^3\left(\frac{1}{3} - \frac{1}{4}\right) + \cdots$$

$$= \left(z + \frac{z^2}{2} + \frac{z^3}{3} + \cdots\right) - \frac{z}{2} - \frac{z^2}{3} - \frac{z^4}{4} \ldots$$

$$= -\log(1 - z) - \frac{1}{z}\left(\frac{z^2}{2} + \frac{z^3}{3} + \frac{z^4}{4} + \cdots\right)$$

$$= -\,log(1-z) + 1 + \frac{1}{z}\,log(1-z)\,, |z| < 1$$

$$= 1 + \left(\frac{1}{z} - 1\right) log(1-z)\,,\ \ |z| < 1$$

$$M_X(t) = \begin{cases} 1 + (e^{-\,t} - 1)\,log(1 - e^{t}) & , \qquad\qquad t < 0 \\ 1 & , \qquad for \quad t = 0 \end{cases}$$

and $M_X(t)$ does not exist for $t > 0$.

3.3

Characteristic Function

In some cases m.g.f. does not exist. For example, consider the p.m.f. given by

$$p(x) = \begin{cases} \frac{6}{\pi^2 x^2}, & x = 1, 2, 3, \ldots \\ 0, & otherwise \end{cases}$$

Its m.g.f. is given by

$$M_X(t) = \sum_{x=1}^{\infty} e^{tx} p(x) = \frac{6}{\pi^2} \sum_{x=1}^{\infty} \frac{e^{tx}}{X^2},$$

which is divergent. Thus, $M_X(t)$ does not exist. A more serviceable function that the m.g.f. is the **characteristic function.**

Characteristic function: The characteristic function of a.r.v. X is defined by

$$\phi_X(t) = E\left[e^{itX}\right] = \begin{cases} \int e^{itx} f(x)dx & if\ X\ is\ a\ c.r.v. with\ p.d.f\ f(x) \\ \sum_x e^{itx}\ p(x) & if\ X\ is\ a\ d.r.v.\ with\ p.m.f\ p(x) \end{cases}$$

where $i = \sqrt{-1}$, the imaginary number.

Note:

1. $|\emptyset_X(t)| = \left|E\left(e^{itX}\right)\right| \le E\left(\left|e^{itX}\right|\right) = E\left(\sqrt{cos^2 tX + sin^2 tX}\right) = E(1) = 1$

   Since $|\emptyset_X(t)| \le 1$, $\emptyset_X(t)$ always exists for any **probability distribution**.

2. $\emptyset_X(t) = E\left[e^{itX}\right] = E\left[1 + (it)X + \frac{(it)^2}{2!}X^2 + \frac{(it)^3}{3!}X^3 + \cdots\right]$

$$= 1 + (it)E(X) + \frac{(it)^2}{2!}E(X^2) + \frac{(it)^3}{3!}E(X^3) + \cdots$$

$$= 1 + (it)\mu_1' + \frac{(it)^2}{2!}\mu_2' + \frac{(it)^3}{3!}\mu_3' + \cdots$$

where $\mu_r' = E(X^r) = r^{th}$ moment abut origin for $r = 1, 2, \ldots$

3. If $\emptyset_X(t)$ is given, then the r^{th} moment about origin is given b

$$\boldsymbol{\mu_r' = coefficient\ of\ \frac{(it)^r}{r!}\ \text{in}\ \emptyset_X(t).}$$

Properties:

1. $\boldsymbol{\emptyset_X(0) = 1}$

Proof: $\emptyset_X(t) = E\left[e^{itX}\right] = E(1)$ when $t = 0$

$$= 1$$

Thus, $\emptyset_X(0) = 1$

2. $|\boldsymbol{\emptyset_X(t)}| \leq \mathbf{1}$ for all real $\boldsymbol{t}$.

Proof: $|\emptyset_X(t)| = \left|E\left(e^{itX}\right)\right| \leq E\left(\left|e^{itX}\right|\right) = E\left(\sqrt{cos^2tX + sin^2tX}\right) = E(1) = 1$

$\Longrightarrow |\emptyset_X(t)| \leq 1$ for all real t

3. $\boldsymbol{\emptyset_X(t)}$ continuous function of $\boldsymbol{t}$ in $(-\infty, \infty)$.

Proof: For $h \neq 0$,

$$|\emptyset_X(t+h) - \emptyset_X(t)| = \left|E\left(e^{i(t+h)X}\right) - E\left(e^{itX}\right)\right| = \left|E\left(e^{i(t+h)X} - e^{itX}\right)\right|$$

$$= \left|E\left\{e^{itX}\left(e^{ihX} - 1\right)\right\}\right| \leq E\left(\left|e^{itX}\right|\left|e^{ihX} - 1\right|\right) = E\left(\left|e^{ihX} - 1\right|\right) \longrightarrow 0 \text{ as } h \longrightarrow 0$$

Thus $\lim\limits_{h\to 0} |\emptyset_X(t+h) - \emptyset_X(t)| = 0$

$\Longrightarrow \lim\limits_{h\to 0} \emptyset_X(t+h) = \emptyset_X(t)$

$\Longrightarrow \emptyset_X(t)$ is a continuous function of t in $(-\infty, \infty)$.

4. $\boldsymbol{\emptyset_X(-t) = \overline{\emptyset_X(t)}}$ **,** $\boldsymbol{i.e}$**,** $\boldsymbol{\emptyset_X(-t)}$ **is the complex conjugate of** $\boldsymbol{\emptyset_X(t)}$**.**

Proof: Here $\overline{\emptyset_X(t)} = \overline{E[e^{itX}]} = E[\cos tX - i\sin tX]$

$\Longrightarrow \emptyset_X(-t) = E[\cos(-tX) + i\sin(-tX)] = E[\cos tX - i\sin tX] = \overline{\emptyset_X(t)}$

Thus, $\emptyset_X(-t) = \overline{\emptyset_X(t)}$

5. If the p.d.f. is even $\boldsymbol{i.e.}$**,** $\boldsymbol{f(-x) = f(x)}$**, then the characteristic function is real valued and even function of** $\boldsymbol{t}$**.**

Proof: We know that , $\phi_X(t) = E\left(e^{itX}\right) = \int_{-\infty}^{\infty} e^{itX} f(x)dx$

Let $x = -y \Longrightarrow dx = -dy$. Then

$$\phi_X(t) = \int_{\infty}^{-\infty} e^{-ity} f(-y)(-dy) = \int_{-\infty}^{\infty} e^{-ity} f(y)dy \qquad (\because\ f\ is\ an\ even\ function)$$

$$= E\left[e^{-itX}\right] = \emptyset_X(-t)$$

Thus, $\emptyset_X(t) = \emptyset_X(-t)$

$\Longrightarrow \emptyset_X(t)$ is an even function of t.

Further, $\overline{\emptyset_X(t)} = \emptyset_X(-t)$ (by property 4)

$= \emptyset_X(t)$ (Since $\emptyset_X(t)$ is even function)

Thus, $\emptyset_X(t)$ is real.

6. If X is a r.v. with characteristic function $\emptyset_X(t)$ and $\mu_r' = E(X^r)$ exists, then

$$\boldsymbol{\mu_r' = (-i)^r \frac{d^r}{dt^r}\left(\emptyset_X(t)\right)\Big|_{t=0}}$$

Proof:

$$\frac{d^r}{dt^r}\left(\emptyset_X(t)\right) = \frac{d^r}{dt^r}\left(E\left(e^{itX}\right)\right) = i^r E\left[X^r e^{itX}\right] = i^r E(X^r)$$

Now, $\frac{d^r}{dt^r}\left(\emptyset_X(t)\right)\Big|_{t=0} = i^r E(X^r)$ and $\mu_r' = E(X^r) = \frac{1}{i^r}\frac{d^r}{dt^r}\left(\emptyset_X(t)\right)\Big|_{t=0}$.

Thus, $\mu_r' = (-i)^r \frac{d^r}{dt^r}\left(\emptyset_X(t)\right)\Big|_{t=0}$

7. Effect of change of origin and scale .

Let $U = \frac{X-a}{h}$ where a and h are constants.

Then $\emptyset_U(t) = E\left[e^{itU}\right] = E\left[e^{it\left(\frac{X-a}{h}\right)}\right] = e^{-\frac{ita}{h}} E\left[e^{i\left(\frac{t}{h}\right)X}\right]$

$$\Longrightarrow \ \emptyset_U(t) = e^{-\frac{ita}{h}} \emptyset_X\left(\frac{t}{h}\right)$$

8. If $X_1, X_2, \ldots, X_n$ are independent, then

$$\boldsymbol{\emptyset_{X_1+\cdots+X_n}(t) = \emptyset_{X_1}(t)\,\emptyset_{X_2}(t) \ldots \emptyset_{X_n}(t)}$$

Proof:

$$\emptyset_{X_1+\cdots+X_n}(t) = E\left[e^{it(X_1+X_2+,\ldots+X_n)}\right] = E[e^{itX_1}.e^{itX_2}\ldots e^{itX_n}]$$

$$(\because\ X_1, X_2, \ldots, X_n \ are\ independent)$$

$$= E\left[e^{itX_1}\right].E\left[e^{itX_2}\right]\ldots E\left[e^{itX_n}\right]$$

$$= \emptyset_{X_1}(t).\emptyset_{X_2}(t)\ldots\emptyset_{X_n}(t)$$

$$\Longrightarrow \ \emptyset_{X_1+\cdots+X_n}(t) = \emptyset_{X_1}(t).\emptyset_{X_2}(t)\ldots\emptyset_{X_n}(t)$$

Note: Converse need not be true.

Uniqueness Theorem for Characteristic Functions:

The characteristic function uniquely determines the distribution. That is,
A necessary and sufficient condition for two distributions with p.d.fs $\boldsymbol{f_1}(.)$ and $\boldsymbol{f_2}(.)$ to be identical is that their characteristic function $\emptyset_1(\boldsymbol{t})$ and $\emptyset_2(\boldsymbol{t})$ are identical.

Example 1: If $\boldsymbol{X \sim B(n, p)}$, find its characteristic function and hence obtain its mean and variance.

Solution: Since $X \sim B(n, p)$, its p.m.f. is given by

$$p(x) = \binom{n}{x} p^x q^{n-x} \text{ for } x = 0, 1, 2, \dots, n$$

The characteristic function of X is given by

$$\phi_X(t) = E\left[e^{itX}\right] = \sum_{x=0}^{n} e^{itx} p(x) = \sum_{x=0}^{n} e^{itx} \binom{n}{x} p^x q^{n-x} = \sum_{x=0}^{n} \binom{n}{x} \left(pe^{it}\right)^x q^{n-x}$$

$\Rightarrow \emptyset_X(t) = \left(q + pe^{it}\right)^n$ and $\frac{d}{dt}\left(\emptyset_X(t)\right) = npi\left(q + pe^{it}\right)^{n-1} e^{it}$

The mean of X is given by

$$\mu = E(X) = \mu' = (-i)\frac{d}{dt}\left(\emptyset_X(t)\right)\Big|_{t=0} = (-i)\left[npi\left(q + pe^{it}\right)^{n-1} e^{it}\right]_{t=0}$$

$$= (-i)\, npi = np$$

Now, $\frac{d^2}{dt^2}\left(\emptyset_X(t)\right) = (npi)\frac{d}{dt}\left[\left(q + pe^{it}\right)^{n-1} e^{it}\right]$

$$= (npi)\left[(n-1)\left(q + pe^{it}\right)^{n-2} pie^{2it} + \left(q + pe^{it}\right)^{n-1} ie^{it}\right]$$

Thus, $\mu_2' = (-i)^2 \frac{d^2}{dt^2}\left(\emptyset_X(t)\right)\Big|_{t=0} = (-i)^2 (npi)[(n-1)pi + i]$

$$= (np)[np - p + 1] = np(np + q) = n^2p^2 + npq$$

$$\Rightarrow \mu_2' = n^2p^2 + npq$$

Therefore, the variance of X is given by

$$\sigma^2 = V(X) = \mu_2' - (\mu_1')^2 = n^2p^2 + npq - n^2p^2$$
$$\Longrightarrow \ \sigma^2 = npq.$$

Example 2: If $X \sim P(\lambda)$, find the characteristic function X and hence obtain its mean and variance.

Solution: Since $X \sim P(\lambda)$, its p.m.f. is given by

$$p(x) = \frac{e^{-\lambda}\lambda^x}{x!} \ , x = 0, 1, , 2 \ldots$$

The characteristic function of X is given by

$$\phi_X(t) = E\left[e^{itX}\right] = \sum_{x=0}^{\infty} e^{itx} p(x) = \sum_{x=0}^{\infty} e^{itx} \frac{e^{-\lambda}\lambda^x}{x!}$$

$$= e^{-\lambda} \sum_{x=0}^{\infty} \frac{\left(\lambda e^{it}\right)^x}{x!} = e^{-\lambda} e^{\lambda e^{it}} = e^{\lambda\left(e^{it}-1\right)}$$

$$\Rightarrow \boldsymbol{j_X(t) = e^{\lambda\left(e^{it}-1\right)}}$$

Now, $\frac{d}{dt}\big(\emptyset_X(t)\big) = e^{\lambda(e^{it}-1)}\lambda i e^{it}$

Thus, the mean is given by

$$\mu = \mu_1' = E(X) = (-i)\frac{d}{dt}(\phi_X(t))\Big|_{t=0} = (-i)(\lambda i) = \lambda$$

$$\Longrightarrow \ \mu = \lambda$$

Now, $\frac{d^2}{dt^2}\big(\emptyset_X(t)\big) = (\lambda i)\frac{d}{dt}\left[e^{\lambda(e^{it}-1)}e^{it}\right]$

$$= (\lambda i)\left[e^{\lambda(e^{it}-1)}\lambda i e^{2it} + e^{\lambda(e^{it}-1)} i e^{it}\right]$$

Thus, μ_2' is given by

$$\mu_2' = (-i)^2 \frac{d^2}{dt^2}\left(\emptyset_X(t)\right)\Big|_{t=0} = (-i)^2(\lambda i)(\lambda i + i) = (-i)^2(i^2)\lambda(\lambda+1)$$

$$= \lambda(\lambda+1) = \lambda^2 + \lambda$$

Hence, the variance is given by $\sigma^2 = \mu_2' - (\mu_1')^2 = \lambda^2 + \lambda - \lambda^2 = \lambda \Longrightarrow \sigma^2 = \lambda$

Example 3: If $X \sim N(\mu, \sigma^2)$, find the characteristic function of X and hence obtain its mean and variance.

Solution: Since $X \sim N(\mu, \sigma^2)$, its p.d.f. is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right], -\infty < x, \mu < \infty, \sigma > 0$$

The characteristic function of X is given by

$$\phi_X(t) = E\left[e^{itX}\right] = \int_{-\infty}^{\infty} e^{itx} f(x)\,dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itx} exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right] dx$$

Let $\frac{x-\mu}{\sigma} = z \Longrightarrow x = \mu + \sigma z \Longrightarrow dx = \sigma\, dz$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itx} exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right] dx = \int_{-\infty}^{\infty} e^{it(\mu+\sigma z)} e^{-\frac{1}{2}z^2}\, dz$$

$$= \frac{e^{it\mu}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} exp\left[-\frac{1}{2}\left(z^2 - 2i\sigma zt\right)\right] dz$$

$$= \frac{e^{it\mu}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} exp\left[-\frac{1}{2}\left(z^2 - 2i\sigma zt + i^2\sigma^2t^2 - i^2\sigma^2t^2\right)\right] dz$$

$$= \frac{e^{it\mu - \frac{\sigma^2 t^2}{2}}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} exp\left[-\frac{1}{2}(z - i\sigma t)^2\right] dz$$

Let $z - i\sigma t = u \Rightarrow dz = du$

$$= e^{it\mu - \frac{\sigma^2 t^2}{2}} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{u^2}{2}} du$$

$$= e^{it\mu - \frac{\sigma^2 t^2}{2}} \qquad \left(\because \quad \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{u^2}{2}} du = 1 \right)$$

$\Rightarrow \ \emptyset_X(t) = e^{itu - \frac{\sigma^2 t^2}{2}}$

Now, $\frac{d}{dt}\left(\emptyset_X(t)\right) = e^{it\mu - \frac{\sigma^2 t^2}{2}}(i\mu - \sigma^2 t)$

Then $\mu_1' = (-i)\frac{d}{dt}\left(\emptyset_X(t)\right)\Big|_{t=0} = (-i)\ (i\mu) = \mu$

Thus, Mean $= E(X) = \mu$.

Now, $\frac{d^2}{dt^2}\left(\emptyset_X(t)\right) = e^{it\mu - \frac{1}{2}\sigma^2 t^2}(i\mu - \sigma^2 t)^2 + e^{it\mu - \frac{1}{2}t^2\sigma^2}(-\sigma^2)$

Thus, $\mu_2' = (-i)^2\ \frac{d^2}{dt^2}\left(\emptyset_X(t)\right)\Big|_{t=0} = (-i)^2\ [i^2\mu^2 - \sigma^2] = (-1)(-\mu^2 - \sigma^2)$

$\Rightarrow \ \mu_2' = \mu^2 + \sigma^2$

Hence, the variance is given by

$V(X) = \mu_2' - (\mu_1')^2 = \mu^2 + \sigma^2 - \mu^2 = \sigma^2$

$\Rightarrow$ Variance $= V(X) = \sigma^2$

Finding p.m.f. (or p.d.f.) when characteristic function is known.

If X is a d.r.v. with characteristic function $\emptyset_X(t)$, then $\emptyset_X(t) = \sum P(X = j)e^{itj}$.

First write the characteristic function in this form and then identify the $P(X = j)$ which is the p.m.f. of the d.r.v. X.

Example 4: Find the p.m.f. of the d.r.v. X whose characteristic function is given by $\emptyset_X(t) = \left(q + pe^{it}\right)^n$.

Solution: We have, $\emptyset_X(t) = \left(q + pe^{it}\right)^n$ and

$$\emptyset_X(t) = \left(q + pe^{it}\right)^n = \sum_{j=0}^{n} \binom{n}{j} \left(pe^{it}\right)^j q^{n-j}$$

$$= \sum_{j=0}^{n} \binom{n}{j} p^j q^{n-j} e^{itj} = \sum_{j=0}^{n} P(X = j) e^{itj}$$

$$= E\left[e^{itX}\right] \text{ where } P(X = j) = \binom{n}{j} p^j q^{n-j}$$

Thus p.m.f. is $p(j) = \binom{n}{j} p^j q^{n-j}$ for $j = 0,1,2, \dots, n$.

Example 5: Find the p.m.f. of a d.r.v. X whose characteristic function is given by

$$\boldsymbol{\emptyset_X(t) = e^{\lambda(e^{it}-1)}}.$$

Solution: We have, $\emptyset_X(t) = e^{\lambda(e^{it}-1)}$

$$\emptyset_X(t) = e^{\lambda(e^{it}-1)} = e^{-\lambda} e^{\lambda e^{it}} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\left(\lambda e^{it}\right)^x}{x!} = \sum_{x=0}^{\infty} \frac{e^{-\lambda}\lambda^x}{x!} e^{itx}$$

$$= \sum_{x=0}^{\infty} P(X = x) e^{itx} = E\left[e^{itX}\right]$$

where $P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$ for $x = 0,1,2, \dots$, which is Poisson distribution with parameter λ.

Theorem 1: If X is a continuous random variable with characteristic function $\emptyset_X(t)$, then its p.d.f. is given by

$$\boldsymbol{f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \emptyset_X(t) dt}$$

Example 6: Find the p.d.f corresponding to the characteristic function

$$\boldsymbol{\emptyset_X(t) = e^{it\mu - \frac{1}{2}t^2\sigma^2}}.$$

Solution:

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\,itx} e^{it\mu - \frac{1}{2}t^2\sigma^2} dt$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\frac{1}{2}[t^2\sigma^2 - 2it(x-\mu)]} dt$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} exp\left[-\frac{1}{2}\left\{t\sigma - i\left(\frac{x-\mu}{\sigma}\right)\right\}^2 + \left(\frac{x-\mu}{\sigma}\right)^2\right] dt$$

$$= \frac{1}{2\pi} exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]\int_{-\infty}^{\infty} exp\left[-\frac{1}{2}\left\{t\sigma - i\left(\frac{x-\mu}{\sigma}\right)\right\}^2\right] dt$$

Let $t\sigma - i\left(\frac{x-\mu}{\sigma}\right) = u \Rightarrow dt = \frac{du}{\sigma}$.

$$= \frac{1}{\sigma\sqrt{2\pi}} exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} exp\left(-\frac{u^2}{2}\right) dt$$

$$= \frac{1}{\sigma\sqrt{2\pi}} exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

Therefore, $f(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right], -\infty < x < \infty$

$\Longrightarrow X \sim N(\mu, \sigma^2)$

Example 7: Find the p.d.f. corresponding to the characteristic function defined by

$$\emptyset(\boldsymbol{t}) = \begin{cases} \boldsymbol{1 - |t|} & , \quad \boldsymbol{|t| \leq 1} \\ \boldsymbol{0} & , \quad |\boldsymbol{t}| > 1 \end{cases}$$

Solution: The p.d.f. of $f(x)$ is given by

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\emptyset(t)dt = \frac{1}{2\pi}\int_{-1}^{1} e^{-itx}\emptyset(t)dt$$

$$= \frac{1}{2\pi}\int_{-1}^{0} e^{-itx}(1+t)dt + \frac{1}{2\pi}\int_{0}^{1} e^{-itx}(1-t)dt$$

($\because$ for $-1 < t < 0, |t| = -t$ and for $0 < t < 1, |t| = t$)

Now,

$$\int_{-1}^{0} e^{-itx}(1+t)dt = \left[\frac{e^{-itx}}{-ix}(1+t)\right]_{-1}^{0} + \frac{1}{ix}\int_{-1}^{0} e^{-itx}dt$$

$$= -\frac{1}{ix} + \frac{1}{ix}\left[\frac{e^{-itx}}{-ix}\right]_{-1}^{0}$$

$$= -\frac{1}{ix} + \frac{1}{(ix)^2}\left(e^{ix} - 1\right)$$

Similarly,

$$\int_{0}^{1} e^{-itx}(1-t)dt = \frac{1}{ix} + \frac{1}{(ix)^2}\left(e^{-ix} - 1\right)$$

$$\therefore f(x) = \frac{1}{2\pi}\left[\frac{1}{(ix)^2}\left\{e^{ix} - 1 + e^{-ix} - 1\right\}\right] = \frac{1}{\pi x^2}\left(1 - \frac{e^{ix} + e^{-ix}}{2}\right)$$

$$\Longrightarrow f(x) = \frac{1}{\pi x^2}(1 - \cos x), -\infty < x < \infty$$

3.4

Cumulant Generating Function

Just as the moment generating function (m.g.f.) $M_X(t)$ or characteristic function (ch.f.) $\emptyset_X(t)$ of a r.v.X *generates its moments*, the logarithm of $M_X(t)$ or $\emptyset_X(t)$ generates a sequence of numbers called the **Cumulants of $\boldsymbol{X}$**. Cumulants are of interest for the following two reasons.

1. Moments in terms of cumulants can be obtained easily when compared to obtaining them from m.g.f. or ch.f.
2. j^{th} cumulant of a sum of independent r.vs is simply the sum of the j^{th} cumulants of the summand.

Since the ch.f. exists for every r.v(the m.g.f. need not exist for some r.vs), the cumulant generating function (c.g.f.) is defined as the logarithm of the ch.f.

Cumulant generating function: Let X be a r.v. with characteristic function $\emptyset_X(t) = E\left[e^{itX}\right]$. The cumulant generating function (c.g.f.) of X is defined by

$$\boldsymbol{K_X(t) = ln\left(\emptyset_X(t)\right)} \qquad \text{...} (1)$$

for all t in some open interval about 0 in $\boldsymbol{R}$, provided the RHS can be expanded as a convergent series in powers of t.

Thus,

$$K_X(t) = k_1(it) + k_2\frac{(it)^2}{2!} + \cdots + k_r\frac{(it)^r}{r!} \qquad \text{....}(2)$$

Note that, $k_j =$ coef of $\frac{(it)^j}{j!}$ in $K_X(t)$ and it is called the $\boldsymbol{j^{th}}$ **Cumulant** of X

We have,

$$\emptyset_X(t) = 1 + \mu_1{}'(it) + \mu_2{}'\frac{(it)^2}{2!} + \cdots + \mu_r{}'\frac{(it)^r}{r!} \qquad \text{....} (3)$$

From $(1), (2)$ and (3), we have

$k_1(it) + k_2 \frac{(it)^2}{2!} + \cdots = ln[1 + \mu_1'(it) + \cdots + \cdots]$

$= \left(\mu_1'(it) + \mu_2' \frac{(it)^2}{2!} + \cdots\right) - \frac{1}{2}\left(\mu_1'(it) + \mu_2' \frac{(it)^2}{2!} + \cdots\right)^2 + \frac{1}{3}\left(\mu_1'(it) + \mu_2' \frac{(it)^2}{2!} + \cdots\right)^3 - \cdots$

$$(\because ln(1 + x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots)$$

Comparing the coefficients of like powers of t, we get the relationship between moments and cumulants. Hence, we have

$k_1 = \mu_1' =$ Mean $= \mu$ and $k_2 = \mu_2' - (\mu_1')^2 =$ variance $= \sigma^2$.
Thus, $\mu = k_1$ and $\sigma^2 = k_2$.

Note:

1. From (2), $K_X(t)$ can be written as

$$K_X(t) = \sum_{j=1}^{\infty} k_j \frac{(it)^j}{j!}$$

   Thus j^{th} cumulant $= k_j =$ coef. of $\frac{(it)^j}{j!}$ in $K_X(t)$.

2. From (2), j^{th} cumulant is obtained as

$$k_j = (-i)^j \left.\frac{d^j K_X(t)}{dt^j}\right|_{t=0}$$

Example 1: If $X \sim B(n, p)$, then obtain the c.g.f. of X and hence obtain its mean and variance.

Solution: Since $X \sim B(n, p)$, its p.m.f. is given by

$$p(x) = \binom{n}{x} p^x q^{n-x}, x = 0,1,2, \dots, n$$

Then the characteristic function of X is given by

$$\emptyset_X(t) = E\left[e^{itX}\right] = \sum_{x=0}^{\infty} e^{itx} \binom{n}{x} p^x q^{n-x} = \sum_{x=0}^{\infty} \binom{n}{x} \left(pe^{it}\right)^x q^{n-x}$$

$$= \left(q + pe^{it}\right)^n$$

$$\Rightarrow \emptyset_X(t) = \left(q + pe^{it}\right)^n$$

Thus, the c.g.f. of X is given by

$$K_X(t) = ln\left(\emptyset_X(t)\right) = ln\left[q + pe^{it}\right]^n$$

$$\Rightarrow K_X(t) = n\, ln\left(q + pe^{it}\right)$$

$$= n\, ln\left[q + p\left(1 + (it) + \frac{(it)^2}{2!} + \cdots\right)\right]$$

$$= n\, ln\left[1 + (it)p + \frac{(it)^2}{2!}p + \cdots\right]$$

$$\Rightarrow K_X(t) = n\left[\left\{(it)p + \frac{(it)^2}{2!}p + \cdots\right\} - \frac{1}{2}\left\{(it)p + \frac{(it)^2}{2!}p + \cdots\right\}^2 + \cdots\right]$$

$$\Rightarrow K_X(t) = (it)(np) + \frac{(it)^2}{2!}(np - np^2) + \cdots$$

$\therefore k_1 =$ coef. of $(it) = np$ and $k_2 =$ coef. of $\frac{(it)^2}{2!} = np - np^2 = np(1-p) = npq$

Thus mean and variance are given by $\mu = k_1 = np$ and $\sigma^2 = k_2 = npq$

Example 2: If $X \sim$ Poisson $P(\lambda)$, then find the c.g.f. of X and hence obtain its mean and variance.

Solution: Since $X \sim P(\lambda)$, , its p.m.f. is given by

$$p(x) = \frac{e^{-\lambda}\lambda^x}{x!}\ , x = 0,1,2, \ldots$$

The characteristic function of X is given by

$$\emptyset_X(t) = E\left[e^{itX}\right] = \sum_{x=0}^{\infty} e^{itx} p(x) = \sum_{x=0}^{\infty} e^{itx} \frac{e^{-\lambda}\lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\left(\lambda e^{it}\right)^x}{x!}$$

$$= e^{-\lambda} e^{\lambda e^{it}} = e^{\lambda\left(e^{it}-1\right)}$$

Thus, $\emptyset_X(t) = e^{\lambda\left(e^{it}-1\right)}$

The c.g.f. of X is given by

$$K_X(t) = ln\left(\emptyset_X(t)\right) = ln\left[e^{\lambda\left(e^{it}-1\right)}\right] = \lambda\left(e^{it} - 1\right)$$

$$= \lambda\left[1 + (it) + \frac{(it)^2}{2!} + \frac{(it)^3}{3!} + \cdots - 1\right]$$

$$\Longrightarrow K_X(t) = (it)\lambda + \frac{(it)^2}{2!}\lambda + \cdots$$

Thus, $k_1 =$ coef. of (it) in $K_X(t) = \lambda$ and $k_2 =$ coef. of $\frac{(it)^2}{2!}$ in $K_X(t) = \lambda$

Hence, mean $=$ variance $= \lambda$

Example 3: If $X \sim NB(r, p)$, then find the c.g.f. of X and hence obtain its mean and variance.

Solution: Since $X \sim NB(r, p)$, its p.m.f is given by

$$p(x) = \binom{-r}{x} p^r (-q)^x, x = 0,1,2, \ldots$$

The characteristic function of X is given by

$$\emptyset_X(t) = E\left[e^{itX}\right] = \sum_{x=0}^{\infty} e^{itx} p(x) = \sum_{x=0}^{\infty} e^{itx} \binom{-r}{x} p^r (-q)^x$$

$$= p^r \sum_{x=0}^{\infty} \binom{-r}{x} \left(-qe^{it}\right)^x = p^r \left(1 - qe^{it}\right)^{-r}$$

$$\Longrightarrow \emptyset_X(t) = p^r \left(1 - qe^{it}\right)^{-r}$$

The c.g.f. is given by

$$K_X(t) = ln\big(\emptyset_X(t)\big) = ln\big(p^r\left(1 - qe^{it}\right)^{-r}\big)$$

$$= r\, ln\, p - r\, ln\big(1 - qe^{it}\big)$$

Now, $\frac{d}{dt}\big(K_X(t)\big) = (-r)\frac{-iqe^{it}}{1-qe^{it}} = \frac{irqe^{it}}{1-qe^{it}}$

$$\therefore k_1 = (-i)\frac{d}{dt}\big(K_X(t)\big)\Big|_{t=0} = (-i)\frac{(irq)}{1-q} = \frac{rq}{p}$$

And $\frac{d^2}{dt^2}\big(K_X(t)\big) = irq\frac{d}{dt}\left[\frac{e^{it}}{1-qe^{it}}\right] = irq\left[\frac{(1-qe^{it})ie^{it}+e^{it}qie^{it}}{(1-qe^{it})^2}\right]$

$$\therefore k_2 = (-i)^2\frac{d^2}{dt^2}\big(K_X(t)\big)\Big|_{t=0} = (-i)^2(irq)(i)\left[\frac{p+q}{p^2}\right]$$

$$\Rightarrow k_2 = \frac{rq}{p^2}$$

Thus mean and variance are given by

$\mu = k_1 = \frac{rq}{p}$ and $\sigma^2 = k_2 = \frac{rq}{p^2}$ respectively.

Example 4: If $X \sim N(\mu, \sigma^2)$, then obtain the c.g.f. of X and hence find its mean and variance.

Solution: If $X \sim N(\mu, \sigma^2)$, then its characteristic function can be shown that

$$\emptyset_X(t) = e^{it\mu - \frac{1}{2}\sigma^2 t^2} \qquad \text{(recall!)}$$

Hence, the c.g.f. is given by

$$K_X(t) = ln\left[e^{it\mu - \frac{1}{2}\sigma^2 t^2}\right] = it\mu - \frac{1}{2}\sigma^2 t^2$$

$$\Rightarrow K_X(t) = (it)\mu + \frac{(it)^2}{2!}\sigma^2$$

$\therefore k_1 =$ coef. of (it) in $K_X(t) = \mu$ and $k_2 =$ coef. of $\frac{(it)^2}{2!}$ in $K_X(t) = \sigma^2$

Thus, mean $= \mu$ and variance $= \sigma^2$.

Example 5: If $X \sim E(\lambda)$, then obtain the c.g.f. of X and hence obtain its mean and variance.

Solution: Since $X \sim E(\lambda)$, its ch.f. can be shown that

$$\emptyset_X(t) = \frac{\lambda}{\lambda - it} = \left(1 - \frac{it}{\lambda}\right)^{-1} \qquad \text{(recall!)}$$

The c.g.f. of X is given by

$$K_X(t) = ln(\emptyset_X(t) = (-1)\, ln\left(1 - \frac{it}{\lambda}\right) = (-1)\left[-\frac{it}{\lambda} - \frac{1}{2}\left(\frac{it}{\lambda}\right)^2 - \cdots\right]$$

$$\Longrightarrow K_X(t) = \left[(it)\frac{1}{\lambda} + \frac{(it)^2}{2!}\frac{1}{\lambda^2} + \cdots\right]$$

Thus, $k_1 =$ coef. of (it) in $K_X(t) = \frac{1}{\lambda}$ and $k_2 =$ coef. of $\frac{(it)^2}{2!}$ in $K_X(t) = \frac{1}{\lambda^2}$.
Thus, the mean and variance are given by $\mu = \frac{1}{\lambda}$ and $\sigma^2 = \frac{1}{\lambda^2}$ respectively.

Properties of Cumulants: Here we develop some useful properties of cumulants. Let $\boldsymbol{k_n(X)}$ be the $\boldsymbol{n}^{\text{th}}$ **cumulant** of a r.v. X.

Theorem 1: $\boldsymbol{k_n(cX) = c^n k_n(X)}$ for some real constant $\boldsymbol{c}$.

Proof: Consider $\emptyset_{cX}(t) = E\left[e^{itcX}\right] = E\left[e^{i(tc)X}\right] = \emptyset_X(tc)$

$$\Longrightarrow \emptyset_{cX}(t) = \emptyset_X(ct)$$

$$\Longrightarrow ln\, \emptyset_{cX}(t) = ln\, \emptyset_X(ct)$$

$$\Longrightarrow K_{cX}(t) = K_X(ct)$$

Then, $\frac{d^n}{dt^n}\left(K_{cX}(t)\right)\Big|_{t=0} = \frac{d^n}{dt^n}\left(K_X(ct)\right)\Big|_{t=0} = \frac{d^n}{ds^n}\left(K_X(s)\right)\Big|_{s=0} c^n$,where $ct = s$

Therefore, $(-i)^n \frac{d^n}{dt^n}\left(K_{cX}(t)\right)\Big|_{t=0} = (-i)^n c^n \frac{d^n}{dt^n}\left(K_X(t)\right)\Big|_{t=0}$

$\Longrightarrow k_n(cX) = c^n k_n(X)$

Theorem 2: $\boldsymbol{k_n(X+b)} = \begin{cases} \boldsymbol{k_n(X)+b} & , \boldsymbol{if} \quad \boldsymbol{n=1} \\ \boldsymbol{k_n(X)} & , \boldsymbol{if} \quad \boldsymbol{n} > 1 \end{cases}$

Proof: $\emptyset_{X+b}(t) = E\left[e^{it(X+b)}\right] = e^{itb} E\left[e^{itX}\right] = e^{itb} \emptyset_X(t)$

$\Longrightarrow \emptyset_{X+b}(t) = e^{itb} \emptyset_X(t)$

$\Longrightarrow ln[\emptyset_{X+b}(t)] = itb + ln[\emptyset_X(t)]$

$\Longrightarrow K_{X+b}(t) = itb + K_X(t)$

$$\Longrightarrow (-i)^n \frac{d^n}{dt^n}\left(K_{X+b}(t)\right)\Big|_{t=0} = (-i)^n \frac{d^n}{dt^n}\left((itb)\right)\Big|_{t=0} + (-i)^n \frac{d^n}{dt^n}\left(K_X(t)\right)\Big|_{t=0}$$

If $n = 1$, then $K_n(X+b) = b + K_n(X)$.

If $n > 1$, $K_n(X+b) = K_n(X)$.

Theorem3: If $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent random variables and $\boldsymbol{S = X + Y}$, then

$$\boldsymbol{k_n(S) = k_n(X) + k_n(Y).}$$

Proof: Since X and Y are independent,

$$\emptyset_S(t) = \emptyset_X(t)\emptyset_Y(t)$$

$\Longrightarrow ln[\emptyset_S(t)] = ln[\emptyset_X(t)] + ln[\emptyset_Y(t)] \Longrightarrow K_S(t) = K_X(t) + K_Y(t)$

$$\Longrightarrow (-i)^n \frac{d^n}{dt^n}\left(K_S(t)\right)\Big|_{t=0} = (-i)^n \frac{d^n}{dt^n}\left(K_X(t)\right)\Big|_{t=0} + (-i)^n \frac{d^n}{dt^n}\left(K_Y(t)\right)\Big|_{t=0}$$

$\Longrightarrow k_n(S) = k_n(X) + k_n(Y)$

Generalization: If $\boldsymbol{S = X_1 + \cdots + X_m}$ where $\boldsymbol{X_1, X_2, \ldots, X_m}$ are independent random variables, then

$$\boldsymbol{k_n(S) = k_n(X_1) + k_n(X_2) + \cdots + k_n(X_m)}$$

Theorem 4: Let $\mu_j' = E(X^j)$ be the j^{th} moment of X about zero for $j = 1, 2, 3, \dots, n$ where $\mu_0' = 1$. Let $k_1, k_2, \dots, k_n$ be the n cumulants of X. Then

$$\mu_{r+1}' = \sum_{j=0}^{r} \binom{r}{j} \mu_j' k_{(r+1-j)} \qquad \dots.(1)$$

for $r = 0,1, \dots, n-1$.

Proof: For $j = 0,1,2, \dots, n$, we have

$$\mu_j' = \frac{d^j}{dt^j}\left(\emptyset_X(t)\right)\Big|_{t=0} \text{ and } k_j = (-i)^j \frac{d^j}{dt^j}\left(K_X(t)\right)\Big|_{t=0}$$

where $\emptyset_X(t) = E\left[e^{itX}\right]$ and $K_X(t) = ln[\emptyset(t)]$ or equivalently, $\emptyset_X(t) = e^{K_X(t)}$. Differentiating this last identify w.r.t. t gives

$$\emptyset_X'(t) = e^{K_X(t)} K_X'(t) \qquad \dots(2)$$

and evaluating this at $t = 0$ gives $i\mu_1' = ik_1 \Longrightarrow \mu_1' = k_1$ holds for $r = 0$.

Differentiating (2) for r times, it gives

$$\emptyset_X{}^{(r+1)}(t) = \sum_{j=0}^{r} \binom{r}{j} \emptyset_X{}^{(j)}(t) K_X{}^{(r+1-j)}(t)$$

(Use Leibnitz theorem for the n^{th} derivative of the product of two functions)

and evaluating this at $t = 0$ gives

$$\mu_{r+1}' = \sum_{j=0}^{r} \binom{r}{j} \mu_j' k_{(r+1-j)} \quad \text{for } r = 0,1, \dots, n-1$$

Note: Taking $r = 0,1,2,3$ in (1) produces

$$\left.\begin{aligned} \mu_1' &= k_1 \\ \mu_2' &= k_2 + \mu_1' k_1 \\ \mu_3' &= k_3 + 2\mu_1' k_2 + \mu_2' k_1 \\ \mu_4' &= k_4 + 3\mu_1' k_3 + 3\mu_2' k_2 + \mu_3' k_1 \end{aligned}\right\} \qquad \dots..(3)$$

These recursive formulae can be used to calculate the (μ')s efficiently from ks and vice versa.

Let $\mu_j = E\left[\left(X - E(X)\right)^j\right] = E\left[(X - \mu_1')^j\right]$ for $j = 1,2,\ldots$ are unknown as **central moments.**

Then formulae (3) simplify to

$\mu_2 = k_2, \mu_3 = k_3,\ \mu_4 = k_4 + 3{k_2}^2$ and $k_2 = \mu_2$, $k_3 = \mu_3$, $k_4 = \mu_4 - 3{\mu_2}^2$

Note: Mean $= \boldsymbol{\mu} = \boldsymbol{k_1}$ **and variance** $= \boldsymbol{\mu_2} = \boldsymbol{\sigma^2} = \boldsymbol{k_2}$**.**

3.5

Probability Generating Function

Let X be a non-negative integer valued random variable with p.m.f. $p(x) = P(X = x)$. Then the **probability generating function** (p.g.f.) of X is defined by

$$\boldsymbol{G_X(t) = E[t^X] = \sum_{x=0}^{\infty} t^x p(x)}$$

where $-1 \leq t \leq 1$ is a dummy variable.

Advantages:

1. It is easy to compute.
2. Moments and some probabilities can be obtained easily.
3. The p.m.f. can be obtained easily from p.g.f.
4. It is easy to handle with sum of independent r.vs.

Effect of linear transformation of p.g.f:

Theorem 1: Let $\boldsymbol{X}$ be a discrete random variable with p.g.f. $\boldsymbol{G_X(t)}$. Let $\boldsymbol{Y = a + bX}$ where $\boldsymbol{a}$ and $\boldsymbol{b}$ are real constants. Then $\boldsymbol{G_Y(t) = t^a G_X\left(t^b\right)}$

Proof: By the definition of probability generating function, we have, $G_X(t) = E[t^X]$. Then

$$G_Y(t) = E\left[t^{(a+bX)}\right] = E[t^a t^{bX}] = t^a E[(t^b)^X] = t^a G_X(t^b)$$

$$\Longrightarrow G_Y(t) = t^a G_X(t^b)$$

Theorem 2: Additive Property: If X and Y are independent random variables, then for constants a, b, we have

$$\boldsymbol{G_{(aX+bY)}(t) = G_X(t^a) + G_Y(t^b)}$$

Proof: $G_{aX+bY}(t) = E[t^{aX+bY}]$ (by P.g.f.)

$$= E[(t^a)^X(t^b)^Y]$$

$$= E[(t^a)^X]E[(t^b)^Y] \quad (\because X\&Y \text{ are indpendent.})$$

$$= G_X(t^a)G_Y(t^b)$$

Thus, $G_{aX+bY}(t) = G_X(t^a)G_Y(t^b)$

Note: In particular, if $a = b = 1$, then $\boldsymbol{G_{X+Y}(t) = G_X(t)G_Y(t)}$

Generalization: If $X_1, X_2, \dots, X_n$ are independent random variables, then

$$\boldsymbol{G_{(X_1+\cdots+X_n)}(t) = G_{X_1}(t)G_{X_2}(t)\dots G_{X_n}(t)}$$

Relationship between p.g.f. and m.g.f :

The p.g.f. and m.g.f. of a random variable X are defined by $G_X(t) = E[t^X]$ and $M_X(t) = E[e^{tX}]$ respectively.

Now, $M_X(t) = E[e^{tX}] = E[(e^t)^X] = G_X(e^t)$

$$\Longrightarrow \boldsymbol{M_X(t) = G_X(e^t)}$$

Further, $G_X(t) = E[t^X] = E\left[e^{ln(t^X)}\right] = E[e^{X\,ln\,t}] = M_X(ln\,t)$

$$\Longrightarrow \boldsymbol{G_X(t) = M_X(ln\,t)}$$

Theorem 3: p.m.f. from p.g.f : Let $G_X(t)$ be the p.g.f. of a discrete r.v. X that can take the values 0,1,2,…. Then the p.m.f. of X is given by

$$\boldsymbol{p(x) = P(X = x) = \frac{1}{x!}G_X^{(x)}(t)\Big|_{t=0}}$$

Proof: By definition, we have

$$G_X(t) = E(t^X) = \sum_{x=0}^{\infty} t^x p(x)$$

$$= P(X=0)t^0 + P(X=1)t^1 + P(X=2)t^2 + \cdots + P(X=x)t^x + \cdots$$

It can be observed that the coefficient of t^x in $G_X(t)$ is $P(X=x)$. To obtain coefficient of t^x, differentiate $G_X(t)$, x times and substitute $t=0$. Thus,

$$G_X^{(x)}(t) = x(x-1)(x-2)\ldots 2\cdot 1\cdot P(X=x) + (x+1)(x) + \cdots 2\cdot 1\cdot t\cdot P(X=x+1) + \cdots$$

When $t=0$, all terms after the first vanish. Thus,

$$\boldsymbol{P(X=x) = \frac{1}{x!} G_X^{(x)}(t)\Big|_{t=0} = \frac{1}{x!} G_X^{(x)}(0)}$$

Computation of moments using p.g.f:

In the derivation of moments, we use *Taylor's expansion*:

Suppose $f(x)$ has derivatives of all orders at $x=a$. The Taylor's expansion of $f(x)$ at the point $x=a$ is givne by

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^n(a)}{n!}(x-a)^n + \cdots$$

$$\Rightarrow f(x) = \sum_{i=0}^{\infty} \frac{f^i(a)}{i!}(x-a)^i$$

The Taylor's expansion of $f(t) = t^X$ about $t=1$ is given by

$$t^X = 1 + X(t-1) + X(X-1)\frac{(t-1)^2}{2!} + X(X-1)(X-2)\frac{(t-1)^3}{3!} + \cdots$$

$$\Rightarrow G_X(t) = E[t^X]$$

$$= 1 + (t-1)E(X) + \frac{(t-1)^2}{2!}E[X(X-1)] + \frac{(t-1)^3}{3!}E[X(X-1)(X-2)] + \cdots$$

Differentiating (1) w.r.t., t r times and setting $t=1$, we get

$$G_X^{(r)}(t)\Big|_{t=1} = E[X(X-1)\dots(X-r+1)]$$

$$\Rightarrow \boldsymbol{E[X(X-1)\dots(X-r+1)] = G_X^{(r)}(1)} \qquad \dots(2)$$

which is known as $\boldsymbol{r}^{\text{th}}$ **factorial moment of** $\boldsymbol{X}$. Using these, we can find the moments about origin as follows:

If $r = 1$ in (2), we have

$$\boldsymbol{\mu_1{}' = E(X) = G_X^{(1)}(1)}$$

If $r = 2$ in (2), we have

$$E[X(X-1)] = E[X^2 - X] = E(X^2) - E(X) = G_X^{(2)}(1)$$

$$\Rightarrow E(X^2) = G_X^{(2)}(1) + E(X) = G_X^{(2)}(1) + G_X^{(1)}(1)$$

Thus, the second moment about origin is given by

$$\boldsymbol{\mu_2{}' = E(X^2) = G_X^{(2)}(1) + G_X^{(1)}(1)}$$

Similarly, we can find any moment about origin.

Computation of mean and variance using p.g.f:

Theorem 4: If the r.v. X has p.g.f. $\boldsymbol{G_X(t)}$, then the mean and variance of $\boldsymbol{X}$ are given by

$$\boldsymbol{\mu = E(X) = G_X^{(1)}(1)} \textbf{ and}$$

$$\boldsymbol{\sigma^2 = V(X) = G_X^{(2)}(1) + G_X^{(1)}(1) - \left[G_X^{(1)}(1)\right]^2}$$

respectively.

Proof: From the above, we have

$$\mu_1{}' = G_X^{(1)}(1),\ \mu_2{}' = G_X^{(2)}(1) + G_X^{(1)}(1)$$

Thus , the mean $\mu = \mu_1{}' = G_X^{(1)}(1)$ and variance $\sigma^2 = \mu_2{}' - (\mu_1{}')^2$

$$\Longrightarrow \sigma^2 = G_X^{(2)}(1) + G_X^{(1)}(1) - \left(G_X^{(1)}(1)\right)^2$$

Convolution formula:

Theorem 5: If X and Y are independent integer-valued random variables with $\boldsymbol{P(X = x) = p_1(x)}$ and $\boldsymbol{P(Y = y) = p_2(y), x = 0, 1, 2, \ldots}$ and $\boldsymbol{y = 0, 1, 2, \ldots}$, then

$$\boldsymbol{P(X + Y = z) = p(z) = \sum_{x=0}^{z} p_1(x)p_2(z - x)}$$

Proof: We have ,

$$G_X(t) = \sum_{x=0}^{\infty} t^x p_1(x) \;\; and \;\; G_Y(t) = \sum_{y=0}^{\infty} t^y p_2(y)$$

Now, $\qquad G_{X+Y}(t) = G_X(t)G_Y(t) \qquad$ (Since X and Y are independt)

$$= \left(\sum_{x=0}^{\infty} t^x p_1(x)\right)\left(\sum_{y=0}^{\infty} t^y p_2(y)\right)$$

$$\Longrightarrow G_{X+Y}(t) = \sum_{x=0}^{\infty}\sum_{y=0}^{\infty} p_1(x)p_2(y)t^{x+y} \qquad \ldots(1)$$

Let $Z = X + Y$. Then

$$G_Z(t) = E[t^Z] = \sum_{z=0}^{\infty} t^z p(z) \qquad \ldots(2)$$

From (1) and(2), we have

$$\sum_{z=0}^{\infty} t^z p(z) = \sum_{x=0}^{\infty}\sum_{y=0}^{\infty} p_1(x)p_2(y)t^{x+y}$$

$$\Rightarrow \sum_{z=0}^{\infty} t^z p(z) = \sum_{z=0}^{\infty}\left(\sum_{x=0}^{z} p_1(x)p_2(z-x)\right)t^z$$

$$\Rightarrow p(z) = \sum_{x=0}^{z} p_1(x)p_2(z-x) \text{ ,for } z = 0,1,2,\dots$$

Example 1: If $X \sim B(n,p)$, then find the p.g.f. of X and hence obtain its mean and variance.

Solution: Since $X \sim B(n,p)$, its p.m.f. is given by

$$p(x) = \binom{n}{x} p^x q^{n-x}, x = 0,1,\dots,n, \qquad 0 < p < 1, q = 1-p$$

The p.g.f. of X is given by

$$G_X(t) = E[t^X] = \sum_{x=0}^{n} t^x p(x) = \sum_{x=0}^{n} t^x \binom{n}{x} p^x q^{n-x}$$

$$= \sum_{x=0}^{n} \binom{n}{x} (tp)^x q^{n-x} = (q+tp)^n$$

$$\Rightarrow G_X(t) = (q+tp)^n$$

Differentiating both sides w.r.t., t we get

$$G_X^{(1)}(t) = n(q+tp)^{n-1}p$$

$\Rightarrow \mu =$ mean $= \mu_1{}' = G_X^{(1)}(1) = np$ and variance is given by

$$\sigma^2 = G_X^{(2)}(1) + G_X^{(1)}(1) - \left[G_X^{(1)}(1)\right]^2$$

But $G_X^{(2)}(t) = np(n-1)(q+tp)^{n-1}p$

$$\Rightarrow G_X^{(2)}(t) = n(n-1)p^2 = n^2p^2 - np^2$$

Therefore, $\sigma^2 = n^2p^2 - np^2 + np - n^2p^2 = np(1-p) \Rightarrow \sigma^2 = npq$

Example 2: If $X \sim P(\lambda)$, then find the p.g.f. of X and hence obtain its mean and variance.

Solution: Since $X \sim P(\lambda)$, its p.m.f. is given by

$$p(x) = \frac{e^{-\lambda}\lambda^x}{x!}, \; x = 0,1,\dots \text{ and } \lambda > 0$$

The p.g.f. of X is given by

$$G_X(t) = E[t^X] = \sum_{x=0}^{\infty} t^x p(x) = \sum_{x=0}^{\infty} t^x \frac{e^{-\lambda}\lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{(t\lambda)^x}{x!} = e^{-\lambda} e^{t\lambda} = e^{\lambda(t-1)}$$

$\Longrightarrow G_X(t) = e^{\lambda(t-1)}$

Differentiating both sides w.r.t. t, we get

$G_X^{(1)}(t) = e^{\lambda(t-1)}\lambda$ and $G_X^{(2)}(t) = e^{\lambda(t-1)}\lambda^2$

Thus, $G_X^{(1)}(1) = \lambda$ and $G_X^{(2)}(1) = \lambda^2$

Hence, the mean and variance are given by $\mu = G_X^{(1)}(1) = \lambda$
and $\sigma^2 = G_X^{(2)}(1) + G_X^{(1)}(1) - \left[G_X^{(1)}(1)\right]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$ respectively.

Example 3: If $X \sim NB(r, p)$, then find the p.g.f. of X and hence obtain its mean and variance.

Solution: Since $X \sim P(\lambda)$, its p.m.f. is given by

$$p(x) = \binom{-r}{x} p^x (-q)^x \, , x = 0,1,2,\dots$$

The p.g.f. of X is given by

$$G_X(t) = E[t^X] = \sum_{x=0}^{\infty} t^x p(x) = \sum_{x=0}^{\infty} t^x \binom{-r}{x} p^r (-q)^x$$

$$= p^r \sum_{x=0}^{\infty} \binom{-r}{x} (-tq)^x = p^r (1 - tq)^{-r}$$

$\Longrightarrow G_X(t) = p^r(1-tq)^{-r}$

$\Longrightarrow G_X^{(1)}(t) = p^r(-r)(1-tq)^{-(r+1)}(-q) = rqp^r(1-tq)^{-(r+1)}$

$\Longrightarrow G_X^{(2)}(t) = rqp^r\big(-(r+1)\big)(1-tq)^{-(r+2)}(-q) = r(r+1)q^2p^r(1-tq)^{-(r+2)}$

Thus, $G_X^{(1)}(t) = rqp^rp^{-(r+1)} = \frac{rq}{p}$ and

$G_X^{(2)}(t) = r(r+1)q^2p^rp^{-(r+2)} = (r^2+r)\frac{q^2}{p^2}$

$$\Longrightarrow G_X^{(2)}(t) = \frac{r^2q^2}{p^2} + \frac{rq^2}{p^2}$$

Thus, $\mu =$ mean $= G_X^{(1)}(1) = \frac{rq}{p}$ and

$\sigma^2 =$ variance $= G_X^{(2)}(1) + G_X^{(1)}(1) - \left[G_X^{(1)}(1)\right]^2$

$$= \frac{r^2q^2}{p^2} + \frac{rq^2}{p^2} + \frac{rq}{p} - \frac{r^2q^2}{p^2} = \frac{rq}{p^2}(q+p) \Longrightarrow \sigma^2 = \frac{rq}{p^2}$$

Example4: If $X \sim G(p)$, then find the p.g.f. of X and hence obtain its mean and variance.

Solution: Since $X \sim G(p)$, its p.m.f. is given by

$$p(x) = q^xp, \qquad x = 0,1,2,\ldots$$

The p.g.f. of X is given by

$$G_X(t) = E[t^X] = \sum_{x=0}^{\infty} t^xp(x) = \sum_{x=0}^{\infty} t^xq^xp = p\sum_{x=0}^{\infty}(tq)^x = \frac{p}{1-tq}$$

$\Longrightarrow G_X(t) = p(1-tq)^{-1}$

$G_X^{(1)}(t) = p(-1)(1-tq)^{-2}(-q) = pq(1-tq)^{-2}$

$$\Rightarrow G_X^{(1)}(1) = \frac{pq}{p^2} = \frac{q}{p}$$

Now, $G_X^{(2)}(t) = pq(-2)(1-tq)^{-3}(-q) = 2pq^2(1-tq)^{-3}$

$$\Rightarrow G_X^{(2)}(1) = \frac{2pq^2}{p^3} = \frac{2q^2}{p^2}$$

Hence, the mean μ and variance σ^2 of X are given by:

$\mu = G_X^{(1)}(1) = \frac{q}{p}$ and

$$\sigma^2 = G_X^{(2)}(1) + G_X^{(1)}(1) - \left[G_X^{(1)}(1)\right]^2$$

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

Example 5:The j.p.m.f. of (X, Y) is given in the following table. Prove or diprove $G_{x+y}(t) = G_X(t)G_y(t)$ iff X and Y are independent.

X \ Y	0	1	2	Total
0	$\frac{1}{9}$	$\frac{2}{9}$	0	$\frac{1}{3}$
1	0	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{1}{3}$
2	$\frac{2}{9}$	0	$\frac{1}{9}$	$\frac{1}{3}$
Total	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	1

Solution: Since $P(X=1, Y=2) = \frac{2}{9} \neq P(X=1)P(Y=2) = \frac{1}{3}\frac{1}{3} = \frac{1}{9}$, it follows that X and Y are not independent.

Now, $G_X(t) = G_Y(t) = \frac{1}{3}(1+t+t^2)$

Let $Z = X + Y$. Then $Z = 0, 1, 2, 3, 4$. Let $p_i = P(Z = i), i = 0,1,2,3,4$.

$p_0 = P(Z = 0) = P(X + Y = 0) = P(X = 0, Y = 0) = \frac{1}{9}$

$p_1 = P(Z = 1) = P(X + Y = 1) = P(X = 0, Y = 1) + P(X = 1, Y = 0) = \frac{2}{9} + 0 = \frac{2}{9}$

$p_2 = P(Z = 2) = P(X + Y = 2) = P(X = 0, Y = 2) + P(X = 1, Y = 1) + P(X = 2, Y = 0)$

$$= 0 + \frac{1}{9} + \frac{2}{9} = \frac{3}{9}$$

$p_3 = P(Z = 3) = P(X + Y = 3) = P(X = 1, Y = 2) + P(X = 2, Y = 1) = \frac{2}{9} + 0 = \frac{2}{9}$

$p_4 = P(Z = 4) = P(X + Y = 4) = P(X = 2, Y = 2) = \frac{1}{9}$

The p.d.f. of $Z = X + Y$ is given by

$G_{X+Y}(t) = \frac{1}{9} + \frac{2}{9}t + \frac{3}{9}t^2 + \frac{2}{9}t^3 + \frac{1}{9}t^4$

$\Longrightarrow\ G_{X+Y}(t) = \frac{1}{9}(1 + 2t + 3t^2 + 2t^3 + t^4) = \left[\frac{1}{3}(1 + t + t^2)\right]^2$

$\Longrightarrow\ G_{X+Y}(t) = G_X(t)G_Y(t)$ but X and Y are not independent. Thus the statement is disproved.

Example 6: Can $G_X(t) = \frac{2}{1+t}$ be the p.d.f. of .r.v. X? Give reasons.

Solution: We have $G_X(1) = \frac{2}{1+1} = \frac{2}{2} = 1$

Further, $G_X(t) = \frac{2}{1+t} = 2(1 + t)^{-1} = 2(1 - t + t^2 - t^3 + \cdots)$

$$\Longrightarrow G_X(t) = 2\sum_{x=0}^{\infty}(-1)^x t^x$$

Thus, $p(x) = P(X = x) = coef. of\ t^x$ in $G_X(t) = 2(-1)^x$

$\Longrightarrow p(x) = 2(-1)^x\ , x = 0, 1, 2, \ldots$

Note that it takes negative values also. Hence, $G_X(t)$ is not a p.g.f.

Example 7 : A fair die is thrown n times. Let S be the total number of points. Show that $P(S = n + 5) = \binom{n+4}{5}\left(\frac{1}{6}\right)^n$.

Solution: The p.g.f. of a single throw is given by:

$$G_X(t) = \sum_{x=1}^{6} t^x p(x) = \sum_{x=1}^{6} \frac{t^x}{6}$$

$$= \frac{1}{6}(t + t^2 + \cdots + t^6) = \frac{t}{6}(1 + t + \cdots + t^5) = \frac{t}{6}\frac{(1-t^6)}{1-t}$$

$$\Rightarrow\ G_X(t) = \frac{t}{6}(1 - t^6)\,(1 - t)^{-1}$$

Since the n throws are identical and independent,

$$G_S(t) = [G_S(t)]^n = \frac{t^n(1-t^6)^n(1-t)^{-n}}{6^n}$$

$$= \frac{t^n}{6^n}\sum_{j=0}^{n}\binom{n}{j}\left(-t^6\right)^j \sum_{k=0}^{\infty}\binom{n+k-1}{k} t^k$$

$$\Rightarrow G_S(t) = \frac{1}{6^n}\sum_{j=0}^{n}\sum_{k=0}^{\infty}(-1)^j\binom{n}{j}\binom{n+k-1}{k}\, t^{k+6j+n}$$

$$= \sum_{k=0}^{\infty} P(S = k + 6j + n)\, t^{k+6j+n}$$

where,

$$P(S=k+6j+n)=\frac{1}{6^n}\sum_{j=0}^{n}(-1)^j\binom{n}{j}\binom{n+k-1}{k}$$

Now , $P(S=n+5)=P(S=k+6j+n)$ with $j=0$ and $k=5$

$$=\frac{1}{6^n}(-1)^0\binom{n}{0}\binom{n+5-1}{5}=\frac{1}{6^n}\binom{n+4}{5}$$

$$\Longrightarrow P(S=n+5)=\frac{1}{6^n}\binom{n+4}{5}$$

Unit-IV

Order Statistics and Limit Theorems

4.1

Order Statistics

Independent and identically distributed random variables:
We say that $X_1, X_2, \dots, X_n$ are *independent* and *identically distributed* random variables (i.i.d.r.vs) if

$$F_{X_1,\dots,X_n}(x_1, \dots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i) \quad \text{(independent)} \qquad \dots (1)$$

and $F_{X_i}(x) = F(x)\ \forall\ i = 1,2, \dots, n$ (identically distributed) $\quad \dots \quad (2)$

where $F_{X_i}(x)$ is the c.d.f. of X_i for $i = 1,2, \dots, n$ and $F_{X_1,\dots,X_n}(x_1, \dots, x_n)$ is the j.c.d.f. of $X_1, \dots, X_n$.

For continuous random variables, the c.d.fs are replaced with p.d.fs in equations (1) and (2) while for discrete random variables the c.d.fs are replaced with p.m.fs.

Definition: We say that $X_1, X_2, \dots, X_n$ is a random sample from a population with c.d.f. $F(x)$ (or p.d.f. $f(x)$ or p.m.f. $p(x)$) if $X_1, \dots, X_n$ are i.i.d.r.vs with common c.d.f. $F(x)$ (or p.d.f. $f(x)$ or p.m.f. $p(x)$).

Definition: Let $X_1, X_2, \dots, X_n$ be a random sample from a population with c.d.f. $F(x)$. Define

$X_{(1)} =$ smallest of $X_1, X_2, \dots, X_n$

$X_{(2)} =$ second smallest of $X_1, X_2, \dots, X_n$,

.

$X_{(r)} = r^{\text{th}}$ smallest of $X_1, X_2, \dots, X_n$,

.

$$X_{(n)} = \text{largest of } X_1, X_2, \dots, X_n.$$

The ordered values $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ are known as the **order statistics** (o.s) of the n r.vs $X_1, X_2, \dots, X_n$.

Note:

1. o.s are r.vs themselves (as functions of $X_1, \dots, X_n$)
2. o.s. satisfy $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$
3. $X_1, X_2, \dots, X_n$ are i.i.d.r.vs but $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ are neither independent nor identically distributed because of order restriction.

Distributions of $\mathbf{o.s.}$ in continuous case:

Let $X_1, X_2, \dots, X_n$ be a random sample from a continuous population with c.d.f. $F(x)$ and p.d.f. $f(x)$.

Marginal distributions:

1) The c.d.f. and p.d.f. of X_n, the n^{th} o.s. are given by

   $\boldsymbol{F_{X_{(n)}}(x) = [F(x)]^n}$ and $\boldsymbol{f_{X_{(n)}}(x) = n[F(x)]^{n-1}f(x)}$respectively.

2) The c.d.f. and p.d.f. of $X_{(1)}$, the first o.s. are given by

   $F_{X_{(1)}}(x) = 1 - [1 - F(x)]^n$ and $f_{X_{(1)}}(x) = n[1 - F(x)]^{n-1}f(x)$ respectively.

3) The c.d.f. and p.d.f. of $X_{(j)}, 1 \leq j \leq n$, the j^{th} o.s. are given by

$$F_{X_{(j)}}(x) = \sum_{i=j}^{n} \binom{n}{i} [F(x)]^i [1 - F(x)]^{n-i} \text{ and}$$

$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} [F(x)]^{j-1} [1 - F(x)]^{n-j} f(x)$$

   respectively.

Joint distributions

4) For $1 \leq i < j \leq n$, the j.p.d.f. of $X_{(\,i\,)}$ and $X_{(\,j\,)}$ is given by
$f_{X_{(\,i\,)},X_{(\,j\,)}}(u,v)$
$= \frac{n!}{(i-1)!(j-i-1)!(n-j)!}[F(u)]^{i-1}[F(v)-F(u)]^{j-i-1}[1-F(v)]^{n-j}f(u)f(v)$
for $-\infty < u < v < \infty$
5) The j.p.d.f. of k-order statistics $X_{(j_1)}, X_{(j_2)}, \ldots, X_{(j_k)}$ where
$1 \leq r_1 < r_2 < \cdots < r_k \leq n$ and $1 \leq k \leq n$ is for $x_1 \leq x_2 \leq \cdots \leq x_k$ given by
$f_{X_{(j_1)},\ldots,X_{(j_k)}}(x_1, \ldots, x_k)$

$$= \frac{n!}{(j_1-1)!\,(j_2-j_1-1)!\ldots(j_k-j_{k-1}-1)!\,(n-j_k)!} \times$$
$$F^{j_1-1}(x_1)[F(x_2)-F(x_1)]^{j_2-j_1-1}\ldots[F(x_k)-F(x_{k-1})]^{j_k-j_{k-1}-1} \times$$
$$\left[[1-F(x_k)]^{n-j_k}\right] f(x_1)f(x_2)\ldots f(x_k)$$

6) The j.p.d.f. of $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ is given by

$$f_{X_{(1)},\ldots,X_{(n)}}(x_1,\ldots,x_n) = \begin{cases} n!\,f(x_1)\ldots f(x_n) \;, & -\infty < x_1 < \cdots < x_n < \infty \\ \quad 0 \;, & otherwise \end{cases}$$

Distribution of Range: Let us obtain the p.d.f. of the r.v. $R_{ij} = X_{(j)} - X_{(i)}$ for $i < j$. The j.p.d.f. of $X_{(i)}$ and $X_{(j)}$ is given by

$f_{X_{(i)},X_{(j)}}(u,v)$

$$= \frac{n!}{(i-1)!(j-i-1)!(n-j)!}[F(u)]^{i-1}[F(v)-F(u)]^{j-i-1}[1-F(v)]^{n-j}f(u)f(v) \ldots.(1)$$

Let $R_{ij} = X_{(j)} - X_{(i)}$and $X = X_{(i)}$ $\Longrightarrow r_{ij} = v - u$ and $x = u$

$\Longrightarrow u = x$ and $v = r_{ij} + x$

The Jacobian of transformation is given by

$$J = \frac{\partial(u,v)}{\partial(x,r_{ij})} = \begin{vmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial r_{ij}} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial r_{ij}} \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ 1 & 1 \end{vmatrix} = 1 \text{ and } |J| = 1 \qquad \ldots(2)$$

From (1) and (2), the j.p.d.f. of $X_{(i)}$ and R_{ij} is given by

$$f_{X_{(i)},R_{ij}}(x, r_{ij})$$

$$= \frac{n!}{(i-1)!(j-i-1)!(n-j)!}[F(x)]^{i-1}\left[F(x+r_{ij}) - F(x)\right]^{j-i-1} \times$$

$$\left[1 - F(x+r_{ij})\right]^{n-j} f(x) f(x+r_{ij}) \quad \ldots (3)$$

From (2), the m.p.d.f. of R_{ij} is given by

$$f_{R_{ij}}\left(r_{ij}\right) = \int_{-\infty}^{\infty} f_{X_{(i)},R_{ij}}\left(x, r_{ij}\right) dx \qquad \ldots (4)$$

Let $j = n$ and $i = 1$. Then the range is given by $W = X_{(n)} - X_{(1)}$. From (3) and (4), the p.d.f. of W is given by

$$g(w) = n(n-1)\int_{-\infty}^{\infty} [F(x+w) - F(x)]^{n-2} f(x+w) f(x) dx$$

The c.d.f. of w is given by

$$G(w) = P(W \le w) = \int_0^w g(u) du$$

$$= \int_0^w \left(n(n-1)\int_{-\infty}^{\infty} [F(x+u) - F(x)]^{n-2} f(x+u) f(x) dx \right) du$$

$$= n\int_{-\infty}^{\infty} f(x)\left[\int_0^w (n-1) f(x+u)[F(x+u) - F(x)]^{n-2}\right] dx$$

$$\Rightarrow G(w) = n\int_{-\infty}^{\infty} f(x)[F(x+w) - F(x)]^{n-1} dx$$

Example 1: Let X_1, X_2, X_3, X_4 be a random sample of size 4 from uniform $[0, \theta]$ distribution. Find the p.d.f. of $X_{(1)}, X_{(3)}$ and $X_{(4)}$.

Solution: Since each $X \sim U(0, \theta)$, its p.d.f. is given by $f(x) = \frac{1}{\theta}$, $0 < x < \theta$ and its c.d.f is given by

$$F(x) = P(X \leq x) = \int_0^x f(t)dt = \int_0^x \frac{1}{\theta} dt = \left[\frac{t}{\theta}\right]_0^x = \frac{x}{\theta}$$

The p.d.f. of $X_{(1)}$ is given by

$$f_{X_{(1)}}(x) = n[1 - F(x)]^{n-1} f(x) = 4\left(1 - \frac{x}{\theta}\right)^{4-1} \frac{1}{\theta}$$

$$\Longrightarrow f_{X_{(1)}}(x) = \frac{4}{\theta}\left(1 - \frac{x}{\theta}\right)^3, 0 < x < \theta$$

The p.d.f of $X_{(3)}$ is given by

$$f_{X_{(3)}}(x) = \frac{4!}{2!\,1!}\left[\frac{x}{\theta}\right]^2 \left(1 - \frac{x}{\theta}\right)^1 \frac{1}{\theta} = \frac{12x^2(\theta - x)}{\theta^4}$$

$$\Longrightarrow f_{X_{(3)}}(x) = \frac{12x^2(\theta - x)}{\theta^4}, 0 < x < \theta$$

The p.d.f of $X_{(4)}$ is given by

$$f_{X_{(4)}}(x) = n[F(x)]^{n-1} f(x) = 4\left(\frac{x}{\theta}\right)^3 \frac{1}{\theta} = \frac{4x^3}{\theta^4}$$

$$\Longrightarrow f_{X_{(4)}}(x) = \frac{4x^3}{\theta^4}, 0 < x < \theta$$

Example 2: Let $X_1, X_2, \dots, X_n$ be i.i.d.r.v s with common p.d.f .

$$\boldsymbol{f(x) = \begin{cases} 1 & , \quad 0 < x < 1 \\ 0 & , \ otherwise \end{cases}}$$

Find (i) p.d.f. of $X_{(j)}$, $1 \le j \le n$

(ii) j.p.d.f. of $X_{(j)}$ and $X_{(k)}$ for $1 \le j < k \le n$

(iii) p.d.f. of $R = X_{(n)} - X_{(1)}$

Solution: Given

p.d.f: $f(x) = 1 \, , 0 < x < 1$

c.g.f: $F(x) = \int_0^x f(t)dt = x \quad \Longrightarrow F(x) = x, 0 < x < 1$

(i) The pdf of $X_{(j)}$ is given by

$$f_{X_{(j)}}(x_j) = \frac{n!}{(j-1)!(x-j)!} {x_j}^{j-1}(1-x_j)^{n-j} \text{ for } 0 < x_j < 1 \, , 1 \le j \le n$$

(ii) The j.p.d.f. of $X_{(j)}$ and $X_{(k)}$ is given by

$$f_{X_{(j)},X_{(k)}}(x_j, x_k) = \frac{n!}{(j-1)!(k-j-1)!(n-k)!} {x_j}^{j-1}(x_k - x_j)^{k-j-1}(1-x_k)^{n-k} \, ,$$

$$0 < x_j < x_k < 1 \text{ where } 1 \le j < k \le n$$

The j.p.d.f. of $X_{(1)}$ and $X_{(n)}$ is given by

$$f_{X_{(1)},X_{(n)}}(x_1, x_n) = n(n-1)(x_n - x_1)^{n-2} \, , 0 < x_1 < x_n < 1$$

(iii) The p.d.f. of $R = X_{(n)} - X_{(1)}$ is given by

$$g(w) = n(n-1)w^{n-2}(1-w), 0 < w < 1$$

Example 3: Let $X_{(1)}, X_{(2)}, X_{(3)}$ be the o.s. of i.i.d.r.vs X_1, X_2, X_3 with common p.d.f.

$$f(x) = \begin{cases} \beta\, e^{-x\beta}, x > 0, \beta > 0 \\ 0\,, otherwise \end{cases}$$

Let $Y_1 = X_{(3)} - X_{(2)}$ and $Y_2 = X_{(2)}$. Show that Y_1 and Y_2 are independent.

Solution: The c.d.f. is given by $F(x) = \int_0^x f(t)dt = 1 - e^{-x\beta}, x > 0.$

Then the j.p.d.f. of $X_{(2)}$ and $X_{(3)}$ is given by

$$f_{X_{(2)},X_{(3)}}(x,y) = \frac{3!}{1!\,0!\,0!}\left(1 - e^{-\beta x}\right)\beta e^{-\beta x}\beta e^{-\beta y}, 0 < x < y < \infty$$

Here $y_1 = y - x$ and $y_2 = x$

$\Longrightarrow x = y_2$ and $y = y_1 + y_2$

The Jacobian of transformation is given by

$$J = \frac{\partial(x,y)}{\partial(y_1,y_2)} = \begin{vmatrix} \frac{\partial x}{\partial y_1} & \frac{\partial x}{\partial y_2} \\ \frac{\partial y}{\partial y_1} & \frac{\partial y}{\partial y_2} \end{vmatrix} = \begin{vmatrix} 0 & 1 \\ 1 & 1 \end{vmatrix} = -1 \text{ and } |J| = 1$$

The j.p.d.f. of Y_1 and Y_2 is given by

$$f(y_1, y_2) = 3!\,\beta^2\left(1 - e^{-\beta y_2}\right)e^{-\beta y_2}e^{-\beta(y_1+y_2)}\,, 0 < y_1 < \infty\,, 0 < y_2 < \infty \quad .... (1)$$

The m.p.d.f. of Y_2 is given by

$$f_2(y_2) = 3!\,\beta e^{-2\beta y_2}\left(1 - e^{-\beta y_2}\right), 0 < y_2 < \infty \qquad .... (2)$$

and the m.p.d.f. of Y_1 is given by

$$f_1(y_1) = \beta e^{-\beta y_1},\ 0 < y_1 < \infty \qquad .... (3)$$

From (1), (2) and (3), Y_1 and Y_2 are independent.

Example 4: Let $X_1, X_2, \dots, X_n$ be a random samples from a population with continuous density. Show that $Y = min(X_1, X_2, \dots, X_n)$ is exponential with parameter $n\lambda$ iff each X_i is exponential with parameter λ.

Solution: Let X_i be the i.i.d exponential variates with parameter λ and p.d.f.

$$f(x) = \lambda e^{-\lambda x}, x > 0, \lambda > 0$$

and $F(x) = P(X \le x) = \int_0^x f(u)du = \lambda \int_0^x e^{-\lambda u} du = 1 - e^{-\lambda x}$

The distribution function of $Y_1 = min(X_1, \dots, X_n)$ is given by

$$F_{Y_1}(y) = 1 - [1 - F(x)]^n = 1 - \left[1 - 1 + e^{-\lambda x}\right]^n = 1 - e^{-(n\lambda)x}$$

which is the distribution function of exponential distribution with parameter $n\lambda$. Thus, $Y_1 \sim exp(n\lambda)$.

Converse: Let $Y_1 = min(X_1, \dots, X_n) \sim exp(n\lambda)$

Then $P(Y_1 \le y) = 1 - e^{-n\lambda y}$

Now, $P(Y_1 \ge y) = 1 - P(Y_1 \le y) = 1 - \left(1 - e^{-n\lambda y}\right) = e^{-n\lambda y}$
$\Rightarrow P[min(X_1, \dots, X_n) \ge y] = e^{-n\lambda y}$

$\Rightarrow P[X_1 \ge y\,,\ X_2 \ge y\,, \dots.,X_n \ge y] = e^{-n\lambda y}$

$\Rightarrow \prod_{i=1}^{n} P\left(X_i \ge y\right) = e^{-n\lambda y}$ $\qquad (\because \quad Xs\ are\ i.d.d)$

$\Rightarrow P(X_i \ge y) = e^{-\lambda y} \quad \Rightarrow P(X_i \le y) = 1 - e^{-\lambda y}$

which is $exp(\lambda)$ distribution. Thus, X_i's are i.d.d $Exp(\lambda)$.

Example 5: For exponential distribution $\boldsymbol{f(x) = e^{-x}, x \geq 0}$, show that the c.d.f. of $\boldsymbol{X_{(n)}}$ in a random sample of size $\boldsymbol{n}$ is $\boldsymbol{F_n(x) = (1 - e^{-x})^n}$. Hence prove that as $\boldsymbol{n \longrightarrow \infty}$, the c.d.f. of $\boldsymbol{X_n - \ln n}$ tends to the limiting form

$$\boldsymbol{exp(-exp(-x)), -\infty < x < \infty.}$$

Solution: Here $f(x) = e^{-x},\ x \geq 0 \implies F(x) = P(X \leq x) = 1 - e^{-x}$.

The c.d.f of $X_{(n)}$ is given by $F_{X_{(n)}} = [F(x)]^n = (1 - e^{-x})^n$

The c.d.f. of $X_{(n)} - \ln n$ is given by

$$G_n(x) = P\left[X_{(n)} - \ln n \leq x\right] = P\left[X_{(n)} \leq x + \ln n\right]$$

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

$$\implies \quad G_n(x) = \left(1 - \frac{e^{-x}}{n}\right)^n$$

$$\implies \lim_{n \to \infty} G_n(x) = \lim_{n \to \infty} \left(1 - \frac{e^{-x}}{n}\right)^n = e^{-e^{-x}} \qquad \left(\because \lim_{n \to \infty} \left(1 + \frac{x}{n}\right)^n = e^x\right)$$

Distribution of O.S. in discrete case: In discrete case there is no magic formula to compute the distribution of any $X_{(j)}$ or any of the joint distributions. A direct computation is the best course of action.

Let $X_1, X_2, \dots X_n$ be a random sample, from a population with p.m.f. $p(x_i) = P(X = x_i)$ for $i = 1, 2, \dots$

Let $r_i = \sum_{k=1}^{i} p(x_k)$. Then $P\left(X_{(j)} \leq x_i\right) = \sum_{k=j}^{n} \binom{n}{k} r_i^k (1 - r_i)^{n-k}$

$$P\left[X_{(j)} = x_i\right] = \sum_{k=j}^{n} \binom{n}{k} \left[r_i^k (1 - r_i)^{n-k} - r_{i-1}^k (1 - r_{i-1})^{n-k}\right]$$

Example 6: Let $X_1, X_2, \ldots X_n$ are i.i.d.r.vs with common geometric p.m.f. given by

$$p_k = P(X = k) = pq^{k-1}, \;\; k = 1, 2, \ldots, \;\; 0 < p < 1, \; q = 1 - p$$

(i) Find p.m.f. of $X_{(r)}, 1 \leq r \leq n$ and

(ii) Show that X_1 and $X_{(2)} - X_{(1)}$ are independent random variables and $X_{(2)} - X_{(1)}$ has a geometric distribution.

Solution:

(i) For any integer $x \geq 1$ and $r \geq 1$,

$$P[X_{(r)} = x] = P[X_{(r)} \leq x] - P[X_{(r)} \leq (x-1)]$$

Now $P(X_{(r)} \leq x) = P[at\ least\ \ r\ of\ X\ s\ are\ \leq x]$

$$= \sum_{i=1}^{r} \binom{n}{i} [P(X_1 \leq x)]^i [P(X_1 > x)]^{n-i}$$

and $P(X_1 \geq x) = \sum_{k=x}^{\infty} pq^{k-1} = (1-p)^{x-1} = q^{x-1}$

It follows that, $P\left[X_{(r)} = x\right] = \sum_{i=r}^{n} \binom{n}{i} q^{(x-1)(n-i)} \left[q^{n-i}\left(1-q^x\right)^i - \left(1-q^{x-1}\right)^i\right]$

$x = 1, 2, \ldots.$

(ii) Let $n = r = 2$. Then , $P[X_{(2)} = x] = pq^{x-1}\,(pq^{x-1} + 2 - 2q^{x-1})\,, x \geq 1$

Also, for integers $x, y \geq 1$, we have $P[X_{(1)} = x,\; X_{(2)} - X_{(1)} = y]$

$$= P[X_{(1)} = x, X_{(2)} = x + y]$$
$$= P[X_1 = x, X_2 = x + y] + P[X_1 = x + y, X_2 = x]$$
$$= 2pq^{x-1}pq^{x+y-1} = 2pq^{2x-2}pq^y$$
$$= P(X_{(1)} = x)P(X_{(2)} = y)$$

and $P(X_{(1)} = 1,\; X_{(2)} - X_{(1)} = 0) = P(X_{(1)} = X_{(2)} = 1) = p^2$

It follows that $X_{(1)}$ and $X_{(2)} - X_{(1)}$ are independent random variables and, moreover, that $X_{(2)} - X_{(1)}$ a geometric distribution.

4.2

Convergence of Sequence of Random Variables

In this module we investigate convergence properties of sequences of random variables. Throughout this module we assume that $\{X_1, X_2, \ldots\}$ or $\{X_n\}$ is a sequence of r.vs and X is a r.v. We consider ***four different modes of convergence for random variables.***

1. **Almost sure convergence**: It is the **probabilistic version of pointwise convergence** known from elementary real analysis. It is also known as **convergence with probability one.**

   The sequence of r.vs $\{X_n\}$ is said to ***converge almost surely*** to a r.v. X if

   $$P\left(\left\{w : \lim_{n\to\infty} X_n(w) = X(w)\right\}\right) = 1$$

   In this case we write $\boldsymbol{X_n \xrightarrow{a.s} X}$ (or $\boldsymbol{X_n \longrightarrow X}$ **with probability** $\mathbf{1}$).

2. **Convergence in probability:** It is essentially mean that the probability that $|X_n - X|$ exceeds any prescribed strictly positive value, converges to zero. The basic idea behind this type of convergence is that the probability of an *unsual* outcome becomes smaller and smaller as the sequence progresses. The sequence of r.vs $\{X_n\}$ is said to ***converge in probability*** to a r.v. X if

   $$\lim_{n\to\infty} P(\{|X_n - X| > \epsilon\}) = 0$$

   for every $\epsilon > 0$.It is denoted by $X_n \xrightarrow{\boldsymbol{P}} X$.

3. **Convergence in $\boldsymbol{r}^{\text{th}}$ mean:** Let $\{X_n\}$ be a sequence of r.vs such that $E(|X_n|^r) < \infty$ for some $r > 0$. We say that X_n ***converges in the $\boldsymbol{r}^{th}$ mean*** to a r.v. X if $E(|X|^r) < \infty$ and

   $$E(|X_n - X|^r) \longrightarrow 0 \text{ as } n \longrightarrow \infty$$

   and we write $\boldsymbol{X_n \xrightarrow{r} X.}$

If $r = 2$, we call it as **convergence in quadratic mean** and it is denoted by $\boldsymbol{X_n \xrightarrow{q.m} X}$

4. **Convergence in distribution: Convergence in distribution** is very frequently used in practice, most often it arises from the application of the **central limit theorem** (to be discussed in module 4.5). Note that a cumulative distribution function (c.d.f) is briefly called as *distribution function (d.f)* also.

Let $\{F_n\}$ be a sequence of cumulative distribution functions (c.d.fs), if there exists a c.d.f. F such that as $n \to \infty$,

$$F_n(x) \longrightarrow F(x)$$

for all x at which F is continuous, then we say that F_n **converges weakly** to F, and it is denoted by $\boldsymbol{F_n \xrightarrow{w} F}$.

If $\{X_n\}$ is a sequence of r.vs and $\{F_n\}$ is the corresponding sequence of c.d.fs, then we say that $\boldsymbol{X_n}$ ***converges in distribution*** (or **law**) to $\boldsymbol{X}$ if there exists an r.v X with c.d.f. F such that $F_n \xrightarrow{w} F$. We write $X_n \xrightarrow{d} X$ or $X_n \xrightarrow{L} X$.

Note: It is quite possible for a given sequence of c.d.fs to converge to a function that is not a c.d.f.

Example: Let $F_n(x) = \begin{cases} 0, x < n \\ 1, x \geq n \end{cases}$

As $n \to \infty$, $F_n(x) \longrightarrow F(x) = 0$ which is not a c.d.f.

Example 1: **Let $\boldsymbol{X_1, X_2, \ldots, X_n}$ be i.i.d.r.vs with common p.d.f**

$$\boldsymbol{f(x)} = \begin{cases} \boldsymbol{\frac{1}{\theta}} & , \ \boldsymbol{0} < x < \theta, \theta > 0 \\ \boldsymbol{0} & , \ \boldsymbol{otherwise} \end{cases}$$

Let $\boldsymbol{X_{(n)} = max(X_1, \ldots, X_n)}$. Then show that $\boldsymbol{X_{(n)} \xrightarrow{L} X}$, where $\boldsymbol{X}$ is degenerate at $\boldsymbol{x = \theta}$.

(Note: We say that a r.v.X is **degenerate at $\boldsymbol{x = \theta}$** if $P(X = \theta) = 1$)

Solution: Corresponding to p.d.f. $f(x) = \frac{1}{\theta}$, the c.d.f. is given by

$$F(x) = \int_0^x f(t)dt = \frac{1}{\theta}\int_0^x dt = \frac{x}{\theta}$$

$$\Rightarrow F(x) = \begin{cases} 0 & , \quad x < 0 \\ \frac{x}{\theta} & , \quad 0 \le x < \theta \\ 1 & , \quad x \ge \theta \end{cases}$$

Then the c.d.f. of $X_{(n)}$ is given by

$$F_n(x) = [F(x)]^n = \begin{cases} 0 & , \quad x < 0 \\ \left(\frac{x}{\theta}\right)^n & , \quad 0 \le x < \theta \\ 1 & , \quad x \ge \theta \end{cases}$$

We see that as $n \longrightarrow \infty$

$$F_n(x) = F(x) = \begin{cases} 0 \;\; if \;\; x < \theta \\ 1 \;\; if \;\; x \ge \theta \end{cases}$$

which is the d.f. of $P(X = \theta) = 1.\ i.e.,$ X is degenerate at $x = \theta$.

Thus $F_n \xrightarrow{w} F$and hence $X_n \xrightarrow{L} X$.

The following example shows that convergence in distribution does not imply convergence of moments.

Example 2: Let $\boldsymbol{F_n}$ be a sequence of c.d.fs defined by

$$\boldsymbol{F_n(x)} = \begin{cases} \mathbf{0} & , \quad \boldsymbol{x} < 0 \\ \mathbf{1} - \dfrac{\mathbf{1}}{\boldsymbol{n}} & , \quad \mathbf{0} \le \boldsymbol{x} < n \\ \mathbf{1} & , \quad \boldsymbol{x} \ge \boldsymbol{n} \end{cases}$$

Show that $\boldsymbol{X_n} \xrightarrow{L} \boldsymbol{X}$ does not imply $\boldsymbol{E\left({X_n}^k\right)} \longrightarrow \boldsymbol{E\left(X^k\right)}$.

Solution: We see that as $n \longrightarrow \infty$

$$F(x) = \begin{cases} 0 \ , \ x < 0 \\ 1 \ , \ x \geq 0 \end{cases}$$

Note that F_n is the c.d.f. of the r.v. X_n with p.m.f.

$$P(X_n = 0) = 1 - \frac{1}{n} \, , P(X_n = n) = \frac{1}{n}$$

and F is the c.d.f. of the r.v. degenerate at 0 $i.e.$, $P(X = 0) = 1$.

Thus, $F_n \xrightarrow{w} F$ and hence $X_n \xrightarrow{L} X$. We have

$E\left({X_n}^k\right) = 0^k\left(1 - \frac{1}{n}\right) + n^k\left(\frac{1}{n}\right) = n^{k-1}$, where k is a positive integer. Also, $E(X^k) = 0^k 1 = 0$. Hence $E\left({X_n}^k\right) \not\rightarrow E(X^k)$ as $n \longrightarrow \infty$

Therefore, $X_n \xrightarrow{L} X$does not imply $E\left({X_n}^k\right) \longrightarrow E(X^k)$.

The next example shows that weak convergence of distribution of function does not imply the convergence of corresponding p.m.fs or p.d.fs.

Example 3: Let $\{X_n\}$ be a sequence of r.vs with p.m.f.

$$\boldsymbol{f_n(x) = P(X_n = x) = \begin{cases} 1 \ , \ if \quad x = 2 + \frac{1}{n} \\ 0 \ , \quad otherwise \end{cases}}$$

Show that $F_n \xrightarrow{w} F$ does not imply $\boldsymbol{f_n \longrightarrow f}$.

Solution: Note that $f_n(x) \longrightarrow f(x)$ as $n \longrightarrow \infty$, where $f(x) = 0$ for all x.

The c.d.f. of X_n is given by

$$F_n(x) = P(X_n \leq x) = \begin{cases} 0\,, \ x < 2 + \frac{1}{n} \\ 1\,, \ x \geq 2 + \frac{1}{n} \end{cases}$$

which converges to

$$F(x) = \begin{cases} 0 & , \ x < 2 \\ 1 & , \ x \geq 2 \end{cases}$$

at all continuity points of F. Since F is the c.d.f. of a r.v. degenerate at $x = 2$ $i.e.$, $P(X = 2) = 1$

$$i.e\ , f(x) = \begin{cases} 1\ , & x = 2 \\ 0\ , & otherwise \end{cases}$$

Thus, convergence of distribution functions does not imply the convergence of corresponding p.m.fs.

Example 4: Let $\{X_n\}$ be a sequence of r.vs with p.m.f $P(X_n = 1) = \frac{1}{n}$ and $P(X_n = 0) = 1 - \frac{1}{n}$. Then show that $X_n \xrightarrow{P} 0$.

Solution: We have $P(|X_n| > \epsilon) = \begin{cases} P(X_n = 1) = \frac{1}{n}\ , & 0 < \epsilon < 1 \\ 0 & , \quad \epsilon \geq 1 \end{cases}$

It follows that $P(|X_n| > \epsilon) \longrightarrow 0$ as $n \longrightarrow \infty$, and we conclude that $X_n \xrightarrow{P} 0$

Example 5: Let $\{X_n\}$ be a sequence of r.vs defined by

$$\boldsymbol{P(X_n = 0) = 1 - \frac{1}{n},\ P(X_n = 1) = \frac{1}{n}\ , n = 1, 2, \ldots}$$

Show that $X_n \xrightarrow{q.m} X$, where $P(X = 0) = 1$.

Solution: Consider $E(|X_n - 0|^2) = E(|X_n|^2) = E\left({X_n}^2\right) = 0^2\left(1 - \frac{1}{n}\right) + 1^2\left(\frac{1}{n}\right)$

$$= \frac{1}{n} \longrightarrow 0 \text{ as } n \longrightarrow \infty$$

Thus, $X_n \xrightarrow{q.m} X$, where X is degenerate at 0.

Example 6: Let $\{X_n\}$ be a sequence of independent r.vs defined by

$$P(X_n = 0) = 1 - \frac{1}{n} \text{ and } P(X_n = 1) = \frac{1}{n}, \quad n = 1, 2, \ldots$$

Show that $X_n \xrightarrow{q.m} 0$ but $X_n \not\xrightarrow{a.s} 0$

Solution: $E(|X_n - 0|^2) = E(|X_n|^2) = 0^2\left(1 - \frac{1}{n}\right) + 1^2\left(\frac{1}{n}\right) = \frac{1}{n} \longrightarrow 0$ as $n \longrightarrow \infty$

Hence $X_n \xrightarrow{q.m} 0$.

Also, $P(X_n = 0 \ \ for\ every\ m \leq n \leq n_0) = \prod_{n=m}^{n_0}\left(1 - \frac{1}{n}\right) = \frac{m-1}{n_0}$ which converges to zero as $n \longrightarrow \infty$ for all values of m. Thus, $X_n \not\xrightarrow{a.s} 0$

Example 7: Let $\{X_n\}$ be a sequence of independent r.vs defined by

$$P(X_n = 0) = 1 - \frac{1}{n^r} \text{ and } P(X_n = n) = \frac{1}{n^r}, r \geq 2, \ n = 1, 2, \ldots$$

Show that $X_n \xrightarrow{a.s} 0$ but $X_n \not\xrightarrow{r} 0$.

Solution: We have $P(X_n = 0 \ \ for\ \ m \leq n \leq n_0) = \prod_{n=m}^{n_0}\left(1 - \frac{1}{n^r}\right)$

As $n_0 \longrightarrow \infty$, the infinite product converges to some nonzero quantity, which itself converges to 1 as $m \longrightarrow \infty$.

That is, $P\left[\lim_{n\to\infty} X_n = 0\right] = 1$. Therefore $X_n \xrightarrow{a.s} 0$

However, $E(|X - 0|^r) = E(|X|^r) = 0^r\left(1 - \frac{1}{n^r}\right) + n^r \times \frac{1}{n^r} = 1$

and hence $E(|X|^r) = 1$ as $n \longrightarrow \infty$. Therefore, $X_n \not\xrightarrow{r} 0$

Thus, $X_n \xrightarrow{a.s} 0$ but $X_n \not\xrightarrow{r} 0$

A sufficient condition for $a.s.$ convergence:

We state a sufficient condition for the $a.s.$ convergence without proof which is sometimes to verify.

$$X_n \xrightarrow{a.s} X \Leftrightarrow \lim_{n\to\infty} P\left[\bigcup_{m=n}^{\infty} |X_n - X| > \epsilon\right] = 0, \qquad \forall\, \epsilon > 0$$

Example 8: Let $\{X_n\}$ be a sequence of r.vs with $P\left(X_n = \pm\frac{1}{n}\right) = \frac{1}{2}$. Show that $X_n \xrightarrow{r} 0$ and $X_n \xrightarrow{a.s} 0$.

Solution: We have $E(|X_n - 0|^r) = E(|X_n|^r) = \frac{1}{n^r}\left(\frac{1}{2}\right) + \frac{1}{n^r}\left(\frac{1}{2}\right) = \frac{1}{n^r} \longrightarrow 0$ as $n \longrightarrow \infty$ and hence $X_n \xrightarrow{r} 0$. It follows that

$$\bigcup_{j=n}^{\infty}\{|X_j| > \varepsilon\} = \{|X_n| > \varepsilon\}$$

Choosing $n > \frac{1}{\epsilon}$, we see that

$$P\left[\bigcup_{j=n}^{\infty}\{|X_j| > \varepsilon\}\right] = P(\{|X_n| > \varepsilon\}) \le P\left(|X_n| > \frac{1}{n}\right) = 0 \text{ as } n \longrightarrow \infty$$

$$\Rightarrow \lim_{n\to\infty} P\left[\bigcup_{j=n}^{\infty}\{|X_j| > \varepsilon\}\right] = 0 \Longrightarrow X_n \xrightarrow{a.s} 0$$

Implications always valid between modes of convergence

We state the following implications always valid between modes of convergence without proof.

1) $X_n \xrightarrow{r} X \Rightarrow X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{d} X$
2) $X_n \xrightarrow{a.s} X \Rightarrow X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{d} X$

Counter examples to implications among the modes of convergence

1) $X_n \xrightarrow{d} X \not\Rightarrow X_n \xrightarrow{P} X$ (**See P1**)
2) $X_n \xrightarrow{P} X \not\Rightarrow X_n \xrightarrow{r} X$ (**See P2**)
3) $X_n \xrightarrow{P} X \not\Rightarrow X_n \xrightarrow{a.s} X$ (**See P3**)
4) $X_n \xrightarrow{r} X \not\Rightarrow X_n \xrightarrow{a.s} X$
5) $X_n \xrightarrow{a.s} X \not\Rightarrow X_n \xrightarrow{r} X$

The following theorem is known as **Slutsky's Theorem** and is very useful in finding the limiting distribution of certain r.vs. This theorem is stated without proof.

Theorem 1: Slutsky's Theorem: Let $\{X_n, Y_n\}, n = 1, 2, \ldots$ be a sequence of pairs of random variables and let c be a constant. If $X_n \xrightarrow{L} X$ and $Y_n \xrightarrow{P} c$, then

(i) $\boldsymbol{X_n + Y_n \xrightarrow{L} X + c}$
(ii) $\boldsymbol{X_n Y_n \xrightarrow{L} cX}$
(iii) $\boldsymbol{\frac{X_n}{Y_n} \xrightarrow{L} \frac{X}{C}}$ **if $\boldsymbol{c \neq 0}$**

An example presented in **P4** as an application of **Slutsky's theorem.**

4.3

Weak Law of Large Numbers

Let $\{X_n\}$ be a sequence of r.vs and let $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ be the mean of first n r.vs. The weak laws deal with *limits of probabilities involving* $\overline{X}_n$. The strong laws deal with *probabilities involving limits of* $\overline{X}_n$.

Definition of Weak Law of Large Numbers

A sequence $\{X_n\}$ of r.vs is said to satisfy the ***Weak Law of Large Numbers (WLLN)*** if

$$\lim_{n\longrightarrow\infty} P\left[\left|\frac{S_n}{n} - E\left(\frac{S_n}{n}\right)\right| < \epsilon\right] = 1$$

for any $\epsilon > 0$, where $S_n = \sum_{i=1}^{n} X_i$, $i.e.,\ \frac{S_n}{n} \xrightarrow{P} E\left(\frac{S_n}{n}\right)$

Theorem1: Let $\{X_n\}$ be a sequence of r.vs and let $S_n = X_1 + \cdots + X_n$ with $B_n = V(S_n) < \infty$. If $\frac{B_n}{n^2} \longrightarrow 0$ as $n \longrightarrow \infty$, then for any $\epsilon > 0$,

$$\lim_{n\longrightarrow\infty} P\left[\left|\frac{S_n}{n} - E\left(\frac{S_n}{n}\right)\right| < \epsilon\right] = 1$$

$i.e.$, $\{X_n\}$ satisfies WLLN.

Proof: On applying Chebychev's inequality to the variable $\frac{S_n}{n}$, we have

$$P\left[\left|\frac{S_n}{n} - E\left(\frac{S_n}{n}\right)\right| \geq \epsilon\right] \leq \frac{V\left(\frac{S_n}{n}\right)}{\epsilon^2} = \frac{V(S_n)}{n^2\epsilon^2} = \frac{B_n}{n^2\epsilon^2} \longrightarrow 0$$

as $n \longrightarrow \infty$. Thus,

$$\lim_{n\longrightarrow\infty} P\left[\left|\frac{S_n}{n} - E\left(\frac{S_n}{n}\right)\right| \geq \epsilon\right] = 0 \Longrightarrow \lim_{n\longrightarrow\infty} P\left[\left|\frac{S_n}{n} - E\left(\frac{S_n}{n}\right)\right| < \epsilon\right] = 1$$

$\Longrightarrow \{X_n\}$ satisfies WLLN.

Corollary 1: Let $\{X_n\}$ be a sequence of r.vs, $\overline{X_n} = \frac{S_n}{n}$ and $\mu = E\left(\frac{S_n}{n}\right)$.

If $\frac{B_n}{n^2} \longrightarrow 0$ as $n \longrightarrow \infty$, then

$$\lim_{n\to\infty} P[\overline{X_n} \leq k] = \begin{cases} 0 & , if \ \ k < \mu \\ 1 & , if \ \ k > \mu \end{cases}$$

Proof: Since WLLN holds for $\{X_n\}$, we have

$$\lim_{n\to\infty} P[|\overline{X_n} - \mu| < \epsilon] = 1 \Longrightarrow \lim_{n\to\infty} P[|\overline{X_n} - \mu| \geq \epsilon] = 0 \qquad \ldots (1)$$

Since $\{\overline{X_n} \leq \mu - \epsilon\} \subset \{|\overline{X_n} - \mu| \geq \epsilon\}$, we have

$$P(\overline{X_n} \leq \mu - \epsilon) \leq P(|\overline{X_n} - \mu| \geq \epsilon)$$

$$\Longrightarrow \lim_{n\to\infty} P(\overline{X_n} \leq \mu - \epsilon) \leq \lim_{n\to\infty} P(|\overline{X_n} - \mu| \geq \epsilon)$$

$$\Longrightarrow \lim_{n\to\infty} P(\overline{X_n} \leq \mu - \epsilon) = 0$$

$\Longrightarrow \lim_{n\to\infty} P(\overline{X_n} \leq k) = 0$, where $k = \mu - \epsilon$, $i.e., k < \mu$ since $\epsilon > 0$

$$\Longrightarrow \lim_{n\to\infty} P(\overline{X_n} \leq k) = 0 \ \ if \ k < \mu$$

Further, $P(\overline{X_n} \leq \mu + \epsilon) + P(|\overline{X_n} - \mu| > \epsilon) \geq 1$, since the region is larger than sample space covered.

$$\Longrightarrow \lim_{n\to\infty} P(\overline{X_n} \leq \mu + \epsilon) \geq 1 \ \ \left(\because \lim_{n\to\infty} P(|\overline{X_n} - \mu| > \epsilon) = 0\right)$$

$$\Longrightarrow \lim_{n\to\infty} P(\overline{X_n} \leq \mu + \epsilon) = 1$$

$\Longrightarrow \lim_{n\to\infty} P(\overline{X_n} \leq k) = 1$ where $k = \mu + \epsilon \ i.e., k > \mu$ since $\epsilon > 0$

$\Longrightarrow \lim_{n\to\infty} P(\overline{X_n} \leq k) = 1$ if $k > \mu$

Thus, $\lim_{n\to\infty} P(\overline{X_n} \leq k) = \begin{cases} 0 , & k < \mu \\ 1 , & k > \mu \end{cases}$

Variations of the WLLN

The following are some special cases of Theorem1 which are stated without proof.

Theorem 2: (Bernoulli's WLLN)

Let $\{X_n\}$ be a sequence of Bernoulli trials with probability of success equal to p. If S_n is the number of successes in n trials, then

$$\lim_{n\longrightarrow\infty} P\left[\left|\frac{S_n - np}{n}\right| < \epsilon\right] = 1,\ \ \forall\, \epsilon > 0$$

Theorem 3: (Khinchine's WLLN)

Let $\{X_n\}$ be a sequence of i.i.d.r.vs with $E(X_i) = \mu < \infty$, $i = 1,2, \dots$, then the WLLNs holds i.e.,

$$\lim_{n\longrightarrow\infty} P\left[\left|\frac{S_n}{n} - \mu\right| > \epsilon\right] = 0$$

Theorem 4: (Bernstein's WLLN)

Let $\{X_n\}$ be a sequence of random variables for which $var(X_n) = {\sigma_n}^2 < k,\ \forall\, i$, where k is independent of n. If $\sigma_{ij} = cov(X_i, X_j) \longrightarrow 0$ as $|i - j| \longrightarrow \infty$ (*Asymptotic uncorrelatedness*) then the WLLN holds.

Example 1: Let $\{X_n\}$ be i.i.d.r.vs with mean μ and variance σ^2, if

$$\frac{{X_1}^2 + {X_2}^2 + \dots + {X_n}^2}{n} \xrightarrow{P} c$$

as $n \longrightarrow \infty$ for some constant $c (0 \leq c < \infty)$, then find c.

Solution: Here $E(X_i) = \mu$ and $V(X_i) = \sigma^2\ \forall\, i$.

Let $S_n = {X_1}^2 + {X_2}^2 + \dots + {X_n}^2$. Then

$E(S_n) = nE\left({X_1}^2\right)$ $\qquad$ ($\because$ Xs are i.i.d.r.vs)

$$= n\left[V(X_1) + \left(E(X_1)\right)^2\right]$$

$$\Rightarrow E(S_n) = n(\sigma^2 + \mu^2)$$

Since $E(X^2) = V(X) + \left(E(X)\right)^2 = \sigma^2 + \mu^2$ exists for each X^2 in S_n, by Khinchine's WLLN, we have

$$\frac{S_n}{n} = \frac{{X_1}^2 + {X_2}^2 + \cdots + {X_n}^2}{n} \quad E\left({X_1}^2\right) = \mu^2 + \sigma^2$$

Thus, $c = \mu^2 + \sigma^2$.

Example 2: If the i.i.d. r.vs $X_k (k = 1, 2, \ldots)$ assume the value $2^{r-2\ln r}$ with probability $\frac{1}{2^r}$, examine if the WLLN holds for the sequence $\{X_k\}$.

Solution:

$$E(X_k) = \sum_{r=1}^{\infty} 2^{r-2\ln r} \cdot \frac{1}{2^r} = \sum_{r=1}^{\infty} \left(2^{-2}\right)^{\ln r} = \sum_{r=1}^{\infty} \left(\frac{1}{4}\right)^{\ln r}$$

$$= \sum_{r=1}^{\infty} (r)^{\ln\left(\frac{1}{4}\right)} \left(\because a^{\ln n} = n^{\ln a}\right)$$

$$= \sum_{r=1}^{\infty} \left(\frac{1}{r}\right)^{\ln 4} \text{ converges since } \ln 4 = 1.39 > 1 \quad \left(\sum_{n=1}^{\infty} \frac{1}{n^p} \text{ converges if } p > 1\right)$$

$$\text{Thus } E(X_k) < \infty$$

Since $\{X_k\}$ are i.i.d.r.vs with $E(X_k) < \infty$, the WLLN holds for the sequence, by Khinchine's theorem.

Example 3: Let $\{X_n\}$ be a sequence of i.i.d $U(0, 1)$ r.vs. For the geometric mean $G_n = (X_1 . X_2 . \ldots . X_n)^{\frac{1}{n}}$, show that $G_n \xrightarrow{P} c$ where c is some constant. Find c.

Solution: Let $Y = -\ln X$ where $X \sim U(0,1)$. The c.d.f. of Y is given by

$$F_Y(y) = P(Y \le y) = P(-\ln X \le y) = P(X \ge e^{-y}) = \int_{e^{-y}}^{1} 1\, dx = 1 - e^{-y}$$

$\Longrightarrow F_Y(y) = 1 - e^{-y}$ and the p.d.f of Y is given by

$f_Y(y) = \frac{d}{dx}\left(F_Y(y)\right) = e^{-y}$ for $y > 0$.

Then $E(Y) = V(Y) = 1$.

Thus, the sequence $\{Y_n\}$ is i.i.d with finite mean $E(Y_n) = 1$. Hence, by Khinchine's WLLN

$$\sum_{i=1}^{n} \frac{Y_i}{n} \xrightarrow{P} E(Y_1) = 1 \qquad \ldots(1)$$

$$But \ \ \ln G_n = \sum_{i=1}^{n} \ln \frac{X_i}{n} = -\sum_{i=1}^{n} \frac{Y_i}{n}$$

$$\Longrightarrow \sum_{i=1}^{n} \frac{Y_i}{n} = -\,ln\, G_n \qquad \ldots(2)$$

From (1) and (2), we have

$$-\,ln\, G_n \xrightarrow{P} 1 \quad i.e., \ \ G_n \xrightarrow{P} e^{-1}$$

Thus, $c = \frac{1}{e}$.

Example 4: Let X_i can have only two values i^α and $-i^\alpha$ with equal probabilities. If $\{X_i\}$ is a sequence of independent r.vs, then show that WLLN holds if $\alpha < \frac{1}{2}$.

Solution: Here $E(X_i) = i^\alpha \frac{1}{2} - i^\alpha \frac{1}{2} = 0$ and

$$V(X_i) = E\left(X_i^{\,2}\right) = i^{2\alpha} \frac{1}{2} + i^{2\alpha} \frac{1}{2} = i^{2\alpha}$$

Let $S_n = \sum\limits_{k=1}^{n} X_k$. Then

$$B_n = V(S_n) = \sum_{i=1}^{n} V(X_i) \qquad (\because X_i\text{s are independent})$$

$$= \sum_{i=1}^{n} i^{2\alpha} = 1^{2\alpha} + 2^{2\alpha} + \cdots + n^{2\alpha}$$

$$= \int_0^n x^{2\alpha} dx \quad (\text{Euler} - \text{Maclaurion formula})$$

$\Longrightarrow B_n = \frac{n^{2\alpha+1}}{2\alpha+1} \Longrightarrow \frac{B_n}{n^2} = \frac{n^{2\alpha+1}}{2\alpha+1} \longrightarrow 0$ as $n \longrightarrow \infty$ if $\alpha < \frac{1}{2}$

Thus, $\frac{B_n}{n^2} \longrightarrow 0$ as $n \longrightarrow \infty$ when $\alpha < \frac{1}{2}$

Therefore, $\{X_n\}$ holds WLLN when $\alpha < \frac{1}{2}$.

4.4

Strong Law of Large Numbers

Definition: A sequence of r.vs $\{X_n\}$ is said to satisfy the **strong law of large numbers** (**SLLN**) if

$$\left[\frac{S_n - E(S_n)}{n}\right] \xrightarrow{a.s} 0 \text{ as } n \longrightarrow \infty$$

We state the following theorems without proof which are useful in checking whether a given *sequence satisfies SLLN or not.*

Theorem1: (Kolmogorov's SLLN)

This theorem is helpful when the r.vs in the sequence are *independent but not identically distributed.*

Statement: Let $\{X_n\}$ be a sequence of independent r.vs with $E(X_i) = \mu$ and $V(X_i) = \sigma_i{}^2 < \infty$ for $= 1, 2, \ldots$. If $\sum_{k=1}^{\infty} \frac{\sigma_k{}^2}{k^2} < \infty$, then the SLLN holds for the sequence $\{X_n\}$.

Theorem 2:

This theorem is helpful when the r.vs in the sequence are independent and identically distributed (i.i.d).

Statement: The sequence $\{X_n\}$ of i.i.d.r.vs holds SLLN iff $E(X_n)$ exists.

Theorem 3: (Borel's SLLN):

This theorem is helpful when the sequence consists of *Bernoulli trials*.

Statement: For a sequence of Bernoulli trials with constant probability of success, the SLLN holds.

Example 1: Let $\{X_n\}$ be a sequence of independent random variables with p.m.f. given by

$$\boldsymbol{P(X_n = \pm 2^n) = \frac{1}{2^{(2n+1)}}\ , P(X_n = 0) = 1 - \frac{1}{2^{2n}}}$$

Does the SLLN hold for $\{X_n\}$?

Solution: We have $E(X_n) = 2^n \frac{1}{2^{2n+1}} - 2^n \frac{1}{2^{2n+1}} = 0$ and

$\sigma_n{}^2 = V(X_n) = E(X_n{}^2) = 2^{2n} \frac{1}{2^{2n+1}} + 2^{2n} \frac{1}{2^{2n+1}} = 1$

Further, $\sum_{n=1}^{\infty} \frac{\sigma_n{}^2}{n^2} = \sum_{n=1}^{\infty} \frac{1}{n^2}$ converges ($\because \sum_{n=1}^{\infty} \frac{1}{n^p}$ converges if $p > 1$).

Hence, the SLLN holds for $\{X_n\}$.

Example 2: For what value of α does the SLLN hold for the sequence

$$\boldsymbol{P(X_k = \pm k^{\alpha}) = \frac{1}{2}}$$

Solution: We have $E(X_k) = k^{\alpha} \frac{1}{2} - k^{\alpha} \frac{1}{2} = 0$ and

$\sigma_k{}^2 = V(X_k) = E(X_k{}^2) = k^{2\alpha} \frac{1}{2} + k^{2\alpha} \frac{1}{2} = k^{2\alpha}$

Further, $\sum_{k=1}^{\infty} \frac{\sigma_k{}^2}{k^2} = \sum_{k=1}^{\infty} \frac{k^{2\alpha}}{k^2} = \sum_{k=1}^{\infty} \frac{1}{k^{2-2\alpha}}$ converges if $2 - 2\alpha > 1$

($\because \sum_{k=1}^{\infty} \frac{1}{k^p}$ converges if $p > 1$).

$\Rightarrow 2\alpha < 1 \Rightarrow \alpha < \frac{1}{2}$

Thus, SLLN holds if $\alpha < \frac{1}{2}$.

Example 3: Let $\{X_n\}$ be a sequence of independent r.vs with p.m.f. given by

$$\boldsymbol{P\left(X_n = \pm\frac{1}{n}\right) = \frac{1}{2}}$$

Check whether SLLN holds for $\{X_n\}$ or not.

Solution: We have $E(X_n) = \frac{1}{n}\frac{1}{2} - \frac{1}{n}\frac{1}{2} = 0$ and

$$\sigma_n{}^2 = V(X_n) = E\left(X_n{}^2\right) = \frac{1}{n^2}\frac{1}{2} + \frac{1}{n^2}\frac{1}{2} = \frac{1}{n^2}$$

Further, $\sum_{n=1}^{\infty}\frac{\sigma_n{}^2}{n^2} = \sum_{n=1}^{\infty}\frac{1}{n^4}$ converges ($\because \sum_{k=1}^{\infty}\frac{1}{k^p}$ converges if $p > 1$).

Therefore, $\{X_n\}$ obeys SLLN.

Example 4: Let $\{X_n\}$ be a sequence of independent r.vs with p.m.f. given by

$$\boldsymbol{P\left(X_k = \pm 2^{-k}\right) = \frac{1}{2}}$$

Check whether SLLN holds or not.

Solution: Here $E(X_k) = 2^{-k}\frac{1}{2} - 2^{-k}\frac{1}{2} = 0$ and

$$\sigma_k{}^2 = V(X_k) = E\left(X_k{}^2\right) = 2^{-2k}\frac{1}{2} + 2^{-2k}\frac{1}{2} = 2^{-2k}$$

Further $\sum_{k=1}^{\infty}\frac{\sigma_k{}^2}{k^2} = \sum_{k=1}^{\infty}2^{-2k}\frac{1}{k^2} = \sum_{k=1}^{\infty}\frac{1}{k^2 2^{2k}}$ converges. Therefore, $\{X_n\}$ obeys the SLLN.

Example 5: Let $\{X_n\}$ be i.i.d.r.vs with mean μ and variance σ^2 and as $n \longrightarrow \infty$,

$$\frac{{X_1}^2 + \cdots + {X_n}^2}{n} \xrightarrow{a.s} c$$

for some constant c $(0 \leq c < \infty)$, then find c.

Solution: Here $E(X_i) = \mu$ and $V(X_i) = \sigma^2 \; \forall \; i$.

Let $S_n = {X_1}^2 + \cdots + {X_n}^2$. Then

$$E(S_n) = nE\left({X_1}^2\right) = n\left[V(X_1) + \left(E(X_1)\right)^2\right] = n(\sigma^2 + \mu^2)$$

$$\Longrightarrow E(S_n) = n(\sigma^2 + \mu^2)$$

$$\Longrightarrow E\left(\frac{S_n}{n}\right) = \sigma^2 + \mu^2$$

By Theorem 2,

$$\frac{S_n}{n} \xrightarrow{a.s} E\left(\frac{S_n}{n}\right) = (\sigma^2 + \mu^2)$$

$$\Longrightarrow \frac{{X_1}^2 + \cdots + {X_n}^2}{n} \xrightarrow{a.s} c, \text{ where } c = \sigma^2 + \mu^2 .$$

Example 6: If the i.i.d.r.vs $\{X_n\}$ assume the value $2^{r-2\ln r}$ with probability $\frac{1}{2^r}$ for $r = 1, 2, \ldots$, examine if the SLLN holds for the sequence $\{X_n\}$.

Solution: By Theorem 2, SLLN holds for i.i.d.r.vs $\{X_n\}$ if $E(X_k)$ exists $\forall \; k$.

Here we have to verify whether $E(X_k)$ is finite or not.

We have

$$E(X_k) = \sum_{r=1}^{\infty} 2^{r-2\ln r} \frac{1}{2^r} = \sum_{r=1}^{\infty} 2^{-2\ln r} = \sum_{r=1}^{\infty} \left(\frac{1}{4}\right)^{\ln r}$$

$$= \sum_{r=1}^{\infty} r^{\ln\left(\frac{1}{4}\right)} \; \left(\because a^{\ln n} = n^{\ln a}\right)$$

$$= \sum_{r=1}^{\infty} \left(\frac{1}{r}\right)^{\ln 4} = \sum_{r=1}^{\infty} \frac{1}{r^{\ln 4}} \text{ where } ln\, 4 = 1.39 > 1$$

which converges.

Thus, $E(X)$ is finite and hence the SLLN holds for $\{X_n\}$.

4.5

Central Limit Theorem

Let $\{X_n\}$ be a sequence of independent random variables. Let $S_n = \sum_{i=1}^{n} X_i$. In laws of large numbers we considered convergence of $\frac{S_n}{n}$ to $E\left(\frac{S_n}{n}\right)$ which is a constant either in *probability* (in case of WLLN) or *almost surely* (in case of SLLN). Here we consider some different situations, namely, $\frac{S_n}{n} \xrightarrow{d} Z$, where Z is a normal variate. If the sequence $\frac{S_n}{n} \xrightarrow{d} Z$, $\frac{S_n}{n}$ is said to follow the **central limit theorem** (CLT) or **normal convergence**. In this module we consider different Central Limit Theorems.

Definition: A sequence of independent r.vs $\{X_i\}$ with mean $E(X_i) = \mu_i$ and $V(X_i) = {\sigma_i}^2 \; \forall \; i$ is said to follow **Central Limit Theorem** (CLT) under certain conditions, if the random variable $S_n = X_1 + X_2 + \cdots + X_n$ is **asymptotically normal (AN)** with mean μ and variance σ^2 where $\mu = \sum_{i=1}^{n} \mu_i$ and $\sigma^2 = \sum_{i=1}^{n} \sigma_i^2$.

Notation: $S_n \sim AN(\mu, \sigma^2)$. Read as S_n follows **asymptotically normal** with mean μ and variance σ^2.

Note:

1. S_n is asymptotically normal means S_n *follows normal distribution as* $n \longrightarrow \infty$.
2. If $Z_n = \frac{(S_n - \mu)}{\sigma}$, then Z_n follows asymptotically standard normal with mean 0 and variance 1 and we write $\boldsymbol{Z_n \sim AN(0, 1)}$.

Variations of the CLT

The following are some variations of the CLT which are stated without proof.

Theorem 1 (De Moivre-Laplace CLT) : If $\{X_n\}$ is a sequence of Bernoulli trials with constant probability of success equal to p, then the distribution of the r.v. $S_n = X_1 + \cdots + X_n$ where X_i's are independent, is asymptotically normal ($i.e$,.

S_n is $AN(np, np(1-p))$

Theorem 2 (Lindeberg-Levy CLT) : This CLT theorem is for i.i.d.r.vs.

If $\{X_i\}$ is a sequence of i.i.d.r.vs with mean $E(X_i) = \mu_1$ and variance $V(X_i) = {\sigma_1}^2$ for all i, then the sum $S_n = X_1 + \cdots + X_n$ is asymptotically normal with mean $\mu = n\mu_1$ and variance $\sigma^2 = n{\sigma_1}^2$.

Theorem 3 (Liapounoff's CLT): This CLT theorem is for independent but not identically distributed random variables.

Let $\{X_i\}$ be a sequence of independent random variables with mean $E(X_i) = \mu_i$ and variance $V(X_i) = {\sigma_i}^2\ \forall\ i$. Let us assume that third absolute moment, say ${\rho_i}^3$ of X_i about its mean exists $i.e.$, ${\rho_i}^3 = E\{|X_i - \mu_i|^3\}$ for $i = 1, 2, \dots, n$ is finite. Let $\rho^3 = \sum_{i=1}^{n} \rho_i^{\,3}, \mu = \sum_{i=1}^{n} \mu_i$ and $\sigma^2 = \sum_{i=1}^{n} \sigma_i^{\,2}$. If $\lim\limits_{n \to \infty} \frac{\rho}{\sigma} = 0$, then the sum $S_n = \sum_{i=1}^{n} X_i$ is $AN(\mu, \sigma^2)$.

Example 1: If $\{X_i\}$ are i.i.d.r.vs with p.m.f $(X_i = \pm 1) = \frac{1}{2}$, find the asymptotic distribution of $S_n = \sum_{i=1}^{n} X_i$.

Solution: Here $E(X_i) = 1.\frac{1}{2} - 1.\frac{1}{2} = 0$ and

$$V(X_i) = E\left({X_i}^2\right) = 1^2.\frac{1}{2} + 1^2.\frac{1}{2} = 1$$

Let $S_n = X_1 + \cdots + X_n$. Then $E(S_n) = E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i) = 0$ and

$V(S_n) = \sum_{i=1}^{n} V(X_i) = \sum_{i=1}^{n} 1 = n$.

Since mean and variance exist for $\{X_i\}$, by Lindeberg-Levy CLT, $S_n \sim AN(0, n)$ or $\frac{S_n}{\sqrt{n}} \sim AN(0,1)$.

Example 2: If $\{X_i\}$ are i.i.d. with $E(X_i) = 0, V(X_i) = \sigma^2$, $0 < \sigma^2 < \infty$ and $\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i$, then show that for any $\epsilon > 0$

$$\boldsymbol{P(\overline{X_n} \geq \epsilon) = \frac{\sigma}{\epsilon\sqrt{n}} \frac{1}{\sqrt{2\pi}} e^{-\frac{n\epsilon^2}{2\sigma^2}}} \textbf{ as } \boldsymbol{n \longrightarrow \infty}.$$

Solution: Let $S_n = X_1 + \cdots + X_n$. Then $E(S_n) = \sum_{i=1}^{n} E(X_i) = 0$ and

$V(S_n) = \sum_{i=1}^{n} V(X_i) = n\sigma^2$. Since $\{X_i\}$ are i.i.d with finite mean and variance, then we have $S_n \sim AN[0, n\sigma^2]$ (by Lindeberg-Levy CLT).

Let $\overline{X_n} = \frac{S_n}{n}$. Then $E(\overline{X_n}) = \frac{1}{n} E(S_n) = \frac{0}{n} = 0$ and

$$V(\overline{X_n}) = \frac{1}{n^2} V(S_n) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

Thus, $\overline{X_n} \sim AN\left[0, \frac{\sigma^2}{n}\right]$.

We have $P(\overline{X_n} \geq \epsilon) = P\left(\frac{\overline{X_n} - 0}{\frac{\sigma}{\sqrt{n}}} \geq \frac{\epsilon - 0}{\frac{\sigma}{\sqrt{n}}}\right) = P\left(Z \geq \frac{\sqrt{n}\epsilon}{\sigma}\right)$ where $Z \sim N(0,1)$

$$= 1 - P\left(Z \leq \frac{\sqrt{n}\epsilon}{\sigma}\right)$$

$$= 1 - \Phi\left(\frac{\sqrt{n}\,\epsilon}{\sigma}\right) \text{, where } \Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{1}{2}t^2}\, dt$$

$$\Longrightarrow P(\overline{X_n} \geq \epsilon) = 1 - \Phi\left(\frac{\sqrt{n}\,\epsilon}{\sigma}\right) \text{ for } \epsilon > 0 \qquad \text{....(1)}$$

But $1 - \Phi(z) = \frac{1}{z}\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}$ ... (2)

(A result in normal distribution)

From (1) and (2), we have

$$P(\overline{X_n} \geq \epsilon) = 1 - \Phi\left(\frac{\sqrt{n}\epsilon}{\sigma}\right) \Longrightarrow P(\overline{X_n} \geq \epsilon) = \frac{\sigma}{\epsilon\sqrt{n}} \cdot \frac{1}{\sqrt{2\pi}}e^{-\frac{n\epsilon^2}{2\sigma^2}}$$

Example 3: Examine if CLT holds for the sequence $\{X_k\}$ with p.m.f $P(X_k = \pm 2^k) = 2^{-(2k+1)}$, $(X_k = 0) = 1 - 2^{-2k}$.

Solution: Since it is a non identically distributed sequence of r.vs, for CLT to hold, we have to verify the Liapounoff's condition.

We have $\mu_k = E(X_k) = 2^k \cdot 2^{-(2k+1)} - 2^k \cdot 2^{-(2k+1)} = 0$,

$$\sigma_k{}^2 = V(X_k) = E(X_k{}^2) = 2^{2k} \cdot 2^{-(2k+1)} + 2^{2k} \cdot 2^{-(2k+1)} = 1 \text{ and}$$

$$\rho_k{}^3 = E\{|X_k - 0|^3\} = E(|X_k|^3) = 2^{3k} \cdot 2^{-(2k+1)} + 2^{3k} \cdot 2^{-(2k+1)}$$

$$= 2 \cdot 2^{3k} \cdot 2^{-(2k+1)} = 2^k$$

Further, we have

$$\mu = \sum_{k=1}^{n} \mu_k = 0\,,$$

$$\sigma^2 = \sum_{k=1}^{n} \sigma_k{}^2 = \sum_{i=1}^{n} 1 = n\,,$$

$$\rho^3 = \sum_{k=1}^{n} \rho_k{}^3 = \sum_{k=1}^{n} 2^k = 2 + 2^2 + \cdots + 2^n = 2(2^n - 1) \text{ and}$$

$$\frac{\rho^3}{(\sigma^2)^{\frac{3}{2}}} = \frac{2(2^n - 1)}{n^{\frac{3}{2}}}$$

Thus, $\lim_{n \to \infty} \frac{\rho^3}{(\sigma^2)^{\frac{3}{2}}} = \lim_{n \to \infty} \frac{2(2^n-1)}{n^{\frac{3}{2}}} = \infty$. Thus, the Liapounoff's condition is not satisfied and hence we cannot say that CLT holds for $\{X_k\}$.

Example 4: Examine if CLT holds for the sequence $\{X_k\}$ with p.m.f

$$\boldsymbol{P(X_k = \pm k^{\alpha}) = \frac{1}{2}.k^{-2\alpha}, P(X_k = 0) = 1 - k^{1-2\alpha}, \alpha < \frac{1}{2}.}$$

Solution: Since it is a non identically distributed of r.vs, for CLT to hold, we have to verify the Liapounov's condition.

We have $\mu_K = E(X_k) = k^{\alpha}.\frac{1}{2}.k^{-2\alpha} - k^{\alpha}.\frac{1}{2}.k^{-2\alpha} = 0$,

$\sigma_k^2 = V(X_k) = E(X_k^2) = k^{2\alpha}.\frac{1}{2}.k^{-2\alpha} + k^{2\alpha}.\frac{1}{2}.k^{-2\alpha} = \frac{1}{2} + \frac{1}{2} = 1$ and

$$\rho_k^3 = E\{|X_k - 0|^3\} = E\{|X_k|^3\} = k^{3\alpha}.\frac{1}{2}.k^{-2\alpha} + k^{3\alpha}.\frac{1}{2}k^{-2\alpha}$$

$$= \frac{1}{2}.k^{\alpha} + \frac{1}{2}k^{\alpha} = k^{\alpha}$$

Further, we have

$$\mu = \sum_{k=1}^{n} \mu_k = \sum_{k=1}^{n} 0 = 0 \ ,$$

$\sigma^2 = \sum_{k=1}^{n} \sigma_k^2 = \sum_{k=1}^{n} 1 = n$ and $\rho^3 = \sum_{k=1}^{n} \rho_k^3 = \sum_{k=1}^{n} k^{\alpha} = 1^{\alpha} + 2^{\alpha} + \ldots + n^{\alpha}$

Note that $\rho^3 \leq n.n^{\alpha} = n^{\alpha+1}$

Now $\lim_{n\to\infty} \frac{\rho^3}{(\sigma^2)^{3/2}} \leq \lim_{n\to\infty} \frac{n^{\alpha+1}}{n^{3/2}} = \lim_{n\to\infty} n^{\alpha - \frac{1}{2}} = 0$, if $\alpha < \frac{1}{2}$

Thus $\lim_{n\to\infty} \frac{\rho^3}{(\sigma^2)^{3/2}} = 0$ if $\alpha < \frac{1}{2}$

Therefore, CLT holds for the sequence $\{X_k\}$.

Applications of central Limit Theorem:

In case of Bernoulli, Binomial and Poisson distributions, evaluation of probabilities using p.m.f. are tedious. Using normal approximation for large samples to these distributions, the probabilities can be easily evaluated.

(a) Let $\{X_n\}$ be a sequence of i.i.d Bernoulli variate $i.e.$, $B(1, p)$.
Let $S_n = X_1 + \cdots + X_n$

Then $S_n \sim B(n, p)$, where $E(S_n) = np$ and $V(S_n) = np(1-p) = npq$

By Lindeberg Levy CLT for large n, $S_n \sim AN\left(E(S_n), V(S_n)\right)$

$$\Rightarrow S_n \sim AN\left(np, np(1-p)\right) \quad \ldots (1)$$

Let $Z_n = \frac{S_n - np}{\sqrt{np(1-p)}}$

Then from (1), $Z_n \xrightarrow{d} Z$ where Z is $N(0,1)$

Thus, $\lim_{n\to\infty} P\left(a \leq \frac{S_n - np}{\sqrt{npq}} \leq b\right) = P(a \leq Z \leq b) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}z^2} dz$

and the RHS can be evaluated using standard normal tables for given real numbers a and b.

(b) Let $\{X_n\}$ be a sequence of i.i.d Binomial variates $.e$, $B(r, p)$.
Let $S_n = X_1 + \cdots + X_n$

Then $E(S_n) = E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} rp = nrp$

$$(\because E(X_i) = rp \ \forall\ i)$$

and $V(S_n) = V\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} V(X_i) \qquad (\because \ X_i s \ are\ independent\ )$

$$= \sum_{i=1}^{n} rp(1-p) = nrp(1-p)$$

$$(\because \quad V(X_i) = rp(1-p))$$

Thus $E(S_n) = nrp$ and $V(S_n) = nrp(1-p)$

By Lindberg – Levy CLT, for large n, we have

$$S_n \sim AN\big(nrp,\ nrp(1-p)\big)$$

Let $Z_n = \frac{S_n - nrp}{\sqrt{nrp(1-p)}}$

Then $Z_n \xrightarrow{d} Z$ where $Z \sim N(0,1)$

Thus, $\lim\limits_{n\to\infty} P\left(a \le \frac{S_n - nrp}{\sqrt{nrp(1-p)}} \le b\right) = P(a \le Z \le b) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}z^2} dz$

and the RHS can be evaluated using standard normal tables for given real numbers a and b.

(c) Let $\{X_n\}$ be a sequence of i.i.d Poisson variates $.e.$, $P(\lambda)$. Let $S_n = \sum_{i=1}^{n} X_i$

Here $E(X_i) = V(X_i) = \lambda \ \forall\ i$. Then $E(S_n) = \sum_{i=1}^{n} E(X_i) = n\lambda$

and $V(S_n) = V\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} V(X_i) = n\lambda$ $\qquad (\because \ X_i\ s\ are\ independent)$

Thus, by Lindeberg – Levy CLT, for large n, $S_n \sim AN(n\lambda, n\lambda)$

Let $Z_n = \frac{S_n - n\lambda}{\sqrt{n\lambda}}$, Then $Z_n \xrightarrow{d} Z$, where $Z \sim N(0,1)$

The probabilities $\lim\limits_{n\to\infty} P\left(a \le \frac{S_n - n\lambda}{\sqrt{n\lambda}} \le b\right) = P(a \le Z \le b) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}z^2} dz$

and the RHS can be evaluated using standard normal for given real numbers a and b.

Example 5: A sample of 100 items is taken at random from a batch known to contain 40% defectives. What is the probability that the sample contains

(i) at least 44 defetives,

(ii) exactly 44 defectives?

Solution:

Let $X_i = \begin{cases} 1 \quad , \quad if\ the\ i^{th}\ item\ is\ defective \\ 0\ , \quad if\ the\ i^{th}\ item\ is\ nondefective \end{cases}$, for $i = 1, 2, \ldots$

It is given that $P(defective) = P(X_i = 1) = 40\% = 0.4$

Then X_i follows Bernoulli distribution $i.e.$, $B(1, 0.4)$

Let $S_n = X_1 + \cdots + X_n$. Then $S_n \sim B(n, p)$

Since $n = 100$ and $p = 0.4$, $S_n \sim B(100, 0.4)$

Since n is large, computation of probabilities using binomial formula is difficult. Hence, by CLT, we use normal approximation to compute the probabilities of S_n instead of binomial distribution.

Here $E(S_n) = np = 100.(0.4) = 40$ and

$$V(S_n) = np(1-p) = 100 \times 0.4 \times 0.6 = 24$$

Let $Z_n = \frac{S_n - E(S_n)}{\sqrt{V(S_n)}} = \frac{S_n - 40}{\sqrt{24}} = \frac{S_n - 40}{4.9}$

Then by Lindeberg Levy CLT, we have $Z_n \xrightarrow{d} Z$, where Z is $N(0, 1)$.

(i) It should be noted that the continuous normal distribution is approximating the discrete binomial distribution so that the continuity correction has to be taken into account in determining the various probabilities. So finding the probability of at least 44 defectives in a sample of 100 items requires finding the area under the normal curve from $\mathbf{43.5}$ **to** $\mathbf{100.5}$

Therefore, the probability of at least 44 defectives is given by

$$P(43.5 < S_n < 100.5) = P\left(\frac{43.5-40}{4.9} < Z < \frac{100.5-40}{4.9}\right)$$

$$= P(0.7143 < Z < 12.347)$$

$$= P(0 < Z < 12.347) - P(0 < Z < 0.7143)$$

$$= 0.5 - 0.2624$$

(See the standard normal distribution table)

$$= 0.2376$$

(ii) The probability of exactly 44 defectives is

$$P(S_n = 44) = P(43.5 < S_n < 44.5)$$

$$= P\left(\frac{43.5-40}{4.9} < Z < \frac{44.5-40}{4.9}\right)$$
$$= P(0.7143 < Z < 0.9184)$$
$$= P(0 < Z < 0.9184) - P(0 < Z < 0.7143)$$
$$= 0.3208 - 0.2624 \quad \text{(See table)}$$
$$= 0.0584$$

Note: Using the binomial distribution, $P(S_n \geq 44) = \sum_{k=44}^{100} \binom{100}{\boldsymbol{k}} (0.4)^k (0.6)^{100-k}$

and $P(S_n = 44) = \binom{100}{44}(0.4)^{44}(0.6)^{56} = 0.0576$ (Using **Binomial tables**).

As can be seen by comparing the answers, both sets of answers are remarkably close.

Example 6: Let $X_1, X_2, \ldots$ be i.i.d. Poisson variables with parameter λ. Use CLT to estimate $P(120 \le S_n \le 160)$,where $S_n = X_1 + \cdots + X_n$, $\lambda = 2$ and $n = 75$

Solution: Since X_i s are i.i.d $P(\lambda), E(X_i) = \lambda = V(X_i)$ for $i = 1, 2, \ldots n$

$\therefore\ E(S_n) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} \lambda = n\lambda$ and

$$V(S_n) = V\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} V(X_i) = n\lambda$$

Hence by Lindberg – Levy CLT, for large n, we have

$S_n \sim AN(n\lambda,\ n\lambda) = AN(150, 150)$. After applying the continuity correction, the required probability is

$p = P(119.5 \le S_n \le 160.5) = P\left(\frac{119.5-150}{\sqrt{150}} \le Z \le \frac{160.5-150}{\sqrt{150}}\right)$,
where $Z \sim N(0, 1)$

$$= P(-2.45 \le Z \le 0.82)$$

$$= P(-2.45 \le Z \le 0) + P(0 \le Z \le 0.82)$$

$$= P(0 \le Z \le 2.45) + P(0 \le Z \le 0.82)$$

$$= 0.4929 + 0.2938 \quad \text{(From standard normal table)}$$

$$= 0.7868$$

Unit-5

Stochastic Processes

5.1

Stationarity of Stochastic Processes

In electrical systems voltage or current waveforms are used as signals for collecting, transmitting or processing information, as well as for controlling and providing power to a variety of devices. These signals (voltage or current waveforms) are functions of time and are of two classes: *deterministic* and *random*. Deterministic signals can be described by the usual mathematical functions with time t as the independent variable. But a random signal always has some element of uncertainty associated with it and hence it is not possible to determine its value exactly at any given point of time. However, we may be able to describe the random signal in terms of its average properties such as the average power in the random signal, its spectral distribution and the probability that the signal amplitude exceeds a given value. The probabilistic model used for characterising a random signal is called a **stochastic process** or **random process**.

A random variable (r.v) is a rule (or function) that assigns a real number to every outcome of a random experiment, while *a stochastic process is a rule* (or *function*) *that assigns a time function to every outcome of a random experiment*.
For example, consider the random experiment of throwing a die at $t = 0$ and observing the number on the top face. The sample space of this experiment consists of the outcomes $\{1,2,3,\dots,6\}$. For each outcome of the experiment, let us arbitrarily assign a function of time $t(0 \le t < \infty)$ in the following manner:

Outcome:	1	2	3	4	5	6
Function of time:	$x_1(t)$	$x_2(t)$	$x_3(t)$	$x_4(t)$	$x_5(t)$	$x_6(t)$

The set of functions $\{x_1(t), x_2(t), \dots, x_6(t)\}$ represents a stochastic process.

Definition: A **stochastic process** is a collection (or ensemble) of r.vs $\{X(s,t)\}$ that are functions of a real variable, namely time t where $s \in S$ (sample space) and $t \in T$ (parameter set or index set).

The set of possible values of any individual member of the stochastic process is called **state space**. Any individual member itself is called a ***sample function*** or a **realization of the process**.

Note:

1. If s and t are fixed, $\{X(s,t)\}$ is a number.
2. If t is fixed $\{X(s,t)\}$ is a r.v.
3. If s is fixed, $\{X(s,t)\}$ is a single time function
4. If s and t are variables, $\{X(s,t)\}$ is a collection of r.vs that are time functions.

Notation: As the dependence of a stochastic process on s is obvious, s will be omitted hereafter in the notation of a stochastic process.
If the parameter set T is discrete, the stochastic process will be noted by $\{X(n)\}$ or $\{X_n\}$.

If the parameter set T is continuous, the process will be denoted by $\{X(t)\}$.

Classification of stochastic Processes

Depending on the continuous or discrete nature of the state space S and parameter set T, a stochastic process can be classified into *four types*:

1. If both T and S are *discrete*, the stochastic process is called a discrete **stochastic sequence**.
   For example, if X_n represents the outcome of the n^{th} throw of a fair die, then $\{X_n, n \geq 1\}$ is a discrete sequence, since $T = \{1,2,3, \dots\}$ and $S = \{1,2,3,4,5,6\}$.
2. If T is discrete and S is continuous, the stochastic process is called a **continuous stochastic sequence.**
   For example, if X_n represents the temperature at the end of the n^{th} hour

of a day, then $\{X_n\ , 1 \leq n \leq 24\}$ is a continuous stochastic sequence, since temperature can take any value in an interval and hence continuous.

3. If T is continuous and S is discrete, the stochastic process is called a **discrete stochastic process.**

   For example, if $X(t)$ represents the number of telephone calls received in the interval $(0, t)$ then $\{X(t)\}$ is a discrete stochastic process, since $S = \{0,1,2,3, \dots\}$.

4. If both T and S are continuous, the stochastic process is called a **continuous stochastic process**.

   For example, if $X(t)$ represents the maximum temperature at a place in the interval $(0, t)$, $\{X(t)\}$ is a continuous stochastic process.

   In the names given above, the word *discrete* or *continuous* is used to refer to the nature of S and the word *sequence* or *process* is used to refer to the nature of T.

Methods of Description of a Stochastic Process

Since a stochastic process is an indexed set of r.vs, we can obviously use the *joint probability distribution functions* to describe a stochastic process.

For a specific $t, X(t)$ is a r.v as was observed earlier.

$F(x, t) = P\{X(t) \leq x\}$ is called the ***first-order distribution*** of the process $\{X(t)\}$ and $f(x, t) = \frac{\partial}{\partial x}\left(F(x, t)\right)$ is called the ***first-order density*** of $\{X(t)\}$.

$F(x_1, x_2, t_1, t_2) = P\{X(t_1) \leq x_1\ ; X(t_2) \leq x_2\}$ is the joint distribution of the r.vs $X(t_1)$ and $X(t_2)$ and is called the ***second-order distribution*** of the process $\{X(t)\}$ and $f(x_1, x_2, t_1, t_2) = \frac{\partial^2}{\partial x_1 \partial x_2} F(x_1, x_2, t_1, t_2)$ is called the ***second-order density*** of $\{X(t)\}$.

Similarly the n^{th} order distribution $\{X(t)\}$ is the joint distribution $F\{x_1, x_2, \dots, x_n\ ; t_1, t_2\ , \dots, t_n\}$ of the r.vs $X(t_1), X(t_2), \dots, X(t_n)$.

The first-order distribution function describes the instantaneous *amplitude* distribution of the process and the second-order distribution function tells us something about the structure of the signal in the time domain and hence the spectral content of the signal. Although the higher-order distributions describe the process in a more detailed manner, the first and second-order distribution functions are primarily used to describe the process.

Special Classes of Stochastic Processes

The important feature of a stochastic process is the relationship among the members of the family. Usually the nature of relationship is understood by the joint distribution function of the member r.vs.
A stochastic process is said to be **specified** only when the parameter set, the state space and the nature of dependence relationship existing among the members of the family are specified.

Based on the dependence relationship among the members of the process, stochastic processes are classified broadly into a few special types such as the ones explained below:

1. **Markov process**

If, for $t_1 < t_2 < t_3 < \cdots < t_n < t$,

$P\{X(t) \leq x \mid X(t_1) = x_1, X(t_2) = x_2, \ldots, X(t_n) = x_n\} = P\{X(t) \leq x \mid X(t_n) = x_n\}$, then the process $\{X(t)\}$ is called a **Markov process**.

In other words, if the future behavior of a process depends only on the present state, but not on the past, the process is a Markov process.

A discrete parameter Markov process is called a **Markov chain**.

2. **Process with independent increments**

If for all choices of $t_1, t_2, \ldots, t_n$ such that $t_1 < t_2 < t_3 < \cdots < t_n$, the random variables $X(t_2) - X(t_1)$, $X(t_3) - X(t_2)$,..., $X(t_n) - X(t_{n-1})$ are independent, then the process $\{X(t)\}$ is said to be a **stochastic process with independent increments.**

Let $T = \{0,1,2,\dots\}$ be the parameter set for $\{X_n\}$. Then $\{Z_n\}$, where $Z_0 = X_0$ and $Z_n = X_n - X_{n-1}$, is a random sequence with independent increments if the r.vs $Z_0, Z_1, Z_2, \dots$, are independent.

Two processes with independent increments play an important role in the theory of random processes. One is the Poisson process that has Poisson distribution for the increments and the other is the Wiener process with a Gaussian distribution for the increments. We will take up the study of Poisson and Gaussian processes in Unit VI.

3. Stationary process

If certain probability distributions or averages do not depend on t, then stochastic process $\{X(t)\}$ is called **stationary**. A rigorous definition and detailed study of stationary processes will be discussed in this module.

Average Values of Stochastic Processes

As in the case of r.vs stochastic processes can be described in terms of averages or expected values, mostly derived from the first and second-order distributions of $\{X(t)\}$. Mean of the process $\{X(t)\}$ is the expected value of a typical member $X(t)$ of the process.

$i.e.,$ $$\mu(t) = E\{X(t)\}$$

Autocorrelation of the process $\{\boldsymbol{X(t)}\}$, denoted by $R_{xx}(t_1, t_2)$ or $R_x(t_1, t_2)$ or $R(t_1, t_2)$, is the expected value of the product of any two members $X(t_1)$ and $X(t_2)$ of the process.

$i.e.,$ $$R(t_1, t_2) = E\{X(t_1) \cdot X(t_2)\}$$

Autocovariance of the process $\{\boldsymbol{X(t)}\}$, denoted by $C_{xx}(t_1, t_2)$ or $C_x(t_1, t_2)$ or $C(t_1, t_2)$, is defined as

$$C(t_1, t_2) = E\big[\big(X(t_1) - \mu(t_1)\big)\big(X(t_2) - \mu(t_2)\big)\big]$$

$$= R(t_1, t_2) - \mu(t_1)\mu(t_2)$$

Correlation co-efficient of the process $\{\boldsymbol{X(t)}\}$, denoted by $\rho_{xx}(t_1, t_2)$ or $\rho(t_1, t_2)$, is defined as

$$\rho(t_1, t_2) = \frac{C(t_1, t_2)}{\sqrt{C(t_1, t_1).C(t_2, t_2)}}$$

where $C(t_1, t_1)$ is the variance of $X(t_1)$.

When we deal with two or more stochastic processes, we can use joint distribution functions or averages to describe the relationship between them.

Cross-correlation of two processes $\{X(t)\}$ and $\{Y(t)\}$ is defined as

$$R_{xy}(t_1, t_2) = E\{X(t_1) \cdot Y(t_2)\}$$

Cross-covariance of two processes $\{X(t)\}$ and $\{Y(t)\}$ is defined as

$$C_{xy}(t_1, t_2) = R_{xy}(t_1, t_2) - \mu_x(t_1)\mu_y(t_2)$$

Cross correlation co-efficient of two processes $\{X(t)\}$ and $\{Y(t)\}$ is defined as

$$\rho_{xy}(t_1, t_2) = \frac{C_{xy}(t_1, t_2)}{\sqrt{C_{xx}(t_1, t_1)C_{yy}(t_2, t_2)}}$$

Stationarity

A stochastic process is called a **strongly stationary process** or **strict sense stationary process** (abbreviated as **SSS** process), if all its finite dimensional distributions are invariant under translation of time parameter.That is, if the joint distribution (and hence the joint density) of $X(t_1), X(t_2), \dots, X(t_n)$ is the same as that of $X(t_1 + h), X(t_2 + h), \dots, X(t_n + h)$ for all $t_1, t_2, \dots, t_n$ and $h > 0$ and for all $n \geq 1$, then the stochastic process $\{X(t)\}$ is called a **SSS process**. If the definition given above holds good for $n = 1,2, \dots, k$ only and not for $n > k$, then the process is called $\boldsymbol{k}^{\text{th}}$ **order stationary**.

Note: If a stochastic process is a SSS process, as per the definition, its first-order densities must be invariant under translation of time, $i.e.$, the densities of $X(t)$ and $X(t + h)$ are the same, $i.e.$, $f(x, t) = f(x, t + h)$.This is possible only if

$f(x, t)$ is independent of t.Therefore, *first-order densities (and hence distribution functions) of a SSS process are independent of time*.

As a consequence, $\mu_r' = E\{X^r(t)\}$*,* $r \geq 1$ *is also independent of time* t*. From this it follows that* $\mu = E\{X(t)\}$ *and variance=* $V\{X(t)\}$ *are independent of time* t

Also the second-order densities must be invariant under translation of time, $i.e.$, the joint p.d.f of $\{X(t_1), X(t_2)\}$ is the same as that of $\{X(t_1 + h), X(t_2 + h)\}$.

$$i.e., f(x_1, x_2, t_1, t_2) = f(x_1, x_2, t_1 + h, t_2 + h).$$

This is possible only if $f(x_1, x_2, t_1, t_2)$ is function of $t = t_1 - t_2$.

Therefore, second-order densities (and hence distribution functions) of a SSS process are functions of $t = t_1 - t_2$.

As a consequence, $R(t_1, t_2) = E\{X(t_1)X(t_2)\}$ is also a function of $t = t_1 - t_2$.It is pointed out that if $E\{X(t)\}$ is a constant and $R(t_1, t_2)$ is a function of $(t_1 - t_2)$, the stochastic process $\{X(t)\}$ need not be a SSS process.

The definition of strict sense Stationarity can be extended as follows:

Two real-valued stochastic processes $\{X(t)\}$ and $\{Y(t)\}$ are said to be **jointly stationary** in the strict sense, if the joint distribution of $X(t)$ and $Y(t)$ are invariant under translation of time.

The complex stochastic process $\{Z(t)\}$, where $Z(t) = X(t) + iY(t)$, is said to be a SSS process if $\{X(t)\}$ and $\{Y(t)\}$ are jointly stationary in the strict sense.

Wide-sense Stationarity

A stochastic process $\{X(t)\}$ with finite first-and second-order moments is called a **weakly stationary process** or **covariance stationary process** or **wide-sense stationary process** (abbreviated as **WSS process**), if its mean is a constant and the autocorrelation depends only on the time difference.

$$i.e., \text{ if } E\{X(t)\} = \mu \text{ and } E\{X(t)X(t - \tau)\} = R(\tau)$$

Note: From the definition given above, it is clear that a SSS process with finite first-and second-order moments is a WSS process, while a WSS process need not be a SSS process.

A stochastic process that is not stationary in any sense is called an ***evolutionary process.***

Two stochastic processes $\{X(t)\}$ and $\{Y(t)\}$ are said to be jointly stationary in the wide sense, if each process is individually a WSS process and $R_{xy}(t_1, t_2)$ is a function of $(t_1 - t_2)$ only.

Example of an SSS Process

Let X_n denote the presence or absence of a pulse at the n^{th} time instant in a digital communication system or digital data processing system.

If $P\{X_n = 1\} = p$ and $P\{X_n = 0\} = 1 - p = q$, then the stochastic processes (sequence) $\{X_n, n \geq 1\}$, called the **Bernoulli's process**, is a SSS process, for its first-order distribution is given by

$X_n = r$	1	0
$P(X_n = r)$	p	q

This distribution is the same for any X_n, $i.e.$, for X_m and X_{m+p}.

Consider the second-order distribution of the process, $i.e.$, the joint distribution of X_r and X_s.

X_s \ X_r	1	0
1	p^2	pq
0	pq	q^2

This joint distribution is the same for the pair of members X_r and X_s and for the pair X_{r+p} and X_{s+p} of the process.

Consider the third-order distribution of the process, $i.e.$, the joint distribution of X_r , X_s and X_t that is given below:

$$P\{X_r = 0, X_s = 0, X_t = 0\} = q^3$$

$$P\{X_r = 0, X_s = 0, X_t = 1\} = pq^2$$

$$P\{X_r = 0, X_s = 1, X_t = 0\} = pq^2$$

$$P\{X_r = 0, X_s = 1, X_t = 1\} = p^2q$$

$$P\{X_r = 1, X_s = 0, X_t = 0\} = pq^2$$

$$P\{X_r = 1, X_s = 0, X_t = 1\} = p^2q$$

$$P\{X_r = 1, X_s = 1, X_t = 0\} = p^2q$$

$$P\{X_r = 1, X_s = 1, X_t = 1\} = p^3$$

This joint distribution is the same for the triple of members X_r, X_s, X_t and for $X_{r+p}, X_{s+p}, X_{t+p}$ of the process and so on, $i.e.$ distributions of all orders are invariant under translation of time.

Note: If $Y_n = \sum_{n=1}^{n} X_n =$ the total number of pulses from time instant 1 through n, then the stochastic processes $\{Y_n , n \geq 1\}$, called the **Binomial process**, is not a SSS process, for $P\{Y_n = i\} = {}^nC_i\, p^i q^{n-i}$ $(i = 0,1,2, \dots , n)$ depends on n, $i.e.$, the distributions of Y_m and Y_{m+p} are not the same).

Analytical Representation of a Stochastic Process

Deterministic signals are usually expressed in simple analytical forms such as $X(t) = e^{-t^2}$ and $Y(t) = 20 \sin 10t$. It is sometimes possible to express a stochastic process in an analytical form using one or more r.vs. For example, consider an FM station that is broadcasting a *tone*, $X(t) = 100 \cos(10^8 t)$, to a large number of receivers distributed randomly in a metropolitan area. The amplitude and phase of the waveform received by any receiver will depend on the distance between the transmitter and the receiver. Since there are a large number of receivers distributed randomly over an area, the distance can be considered as a continuous r.v. Since the amplitude and the phase are functions of distance, they are also r.vs. So we can represent the ensemble (collection) of received waveforms by a stochastic process $\{X(t)\}$ of the form

$$X(t) = A\cos(10^8 t + \theta)$$

where A and θ are r.vs representing the amplitude and phase of the received waveforms.

Such representation of a stochastic process in terms of one or more r.vs whose probability law is known is used in several applications in communication systems.

Example 1: Examine whether Poisson process $\{X(t)\}$. Given by the probability law $P\{X(t) = r\} = \frac{e^{-\lambda t}(\lambda t)^r}{r!}$, $r = 0, 1, 2, \ldots$, is covariance stationary.

Solution: The probability distribution of $X(t)$ is a Poisson distribution with parameter λt.

$\therefore$ $$E\{X(t)\} = \lambda t \neq a \text{ constant.}$$

Therefore, the Poisson process is not covariance stationary.

Example 2:The process $\{X(t)\}$ whose probability distribution is given by

$$\mathrm{P}\{X(t) = n\} = \frac{(at)^{n-1}}{(1+at)^{n+1}}, n = 1, 2, \ldots$$

$$= \frac{at}{1+at}, n = 0$$

Show that it is not stationary.

Solution: The probability distribution of $X(t)$ is

$X(t) = n$	0	1	2	3	...
P_n	$\frac{at}{1+at}$	$\frac{1}{(1+at)^2}$	$\frac{at}{(1+at)^3}$	$\frac{(at)^2}{(1+at)^4}$	...

$$E\{X(t)\} = \sum_{n=0}^{\infty} np_n = \frac{1}{(1+at)^2} + \frac{2at}{(1+at)^3} + \frac{3(at)^2}{(1+at)^4} + \cdots$$

$$= \frac{1}{(1+at)^2}\ \{1 + 2\alpha + 3\alpha^2 + \cdots\}, \text{ where } \alpha = \frac{at}{1+at}$$

$$= \frac{1}{(1+at)^2}(1-\alpha)^{-2} = \frac{1}{(1+at)^2}(1+at)^2 = 1$$

$$E\{X^2(t)\} = \sum_{n=0}^{\infty} n^2 p_n = \sum_{n=1}^{\infty} n^2 \frac{(at)^{n-1}}{(1+at)^{n+1}}$$

$$= \frac{1}{(1+at)^2}\left[\sum_{n=1}^{\infty} n(n+1)\left(\frac{at}{1+at}\right)^{n-1} - \sum_{n=1}^{\infty} n\left(\frac{at}{1+at}\right)^{n-1}\right]$$

$$= \frac{1}{(1+at)^2}\left[\frac{2}{\left(1-\frac{at}{1+at}\right)^3} - \frac{1}{\left(1-\frac{at}{1+at}\right)^2}\right]$$

$$= 1 + 2at$$

$$\therefore Var\{X(t)\} = E\{X^2(t)\} - (E\{X(t)\})^2 = 2at$$

Since $Var\left(X(t)\right)$ is a function of t, the given process is not stationary.

Example 3: Given a r.v Y with characteristic function

$$\boldsymbol{\emptyset(w) = E\left(e^{i\omega Y}\right) = E(\cos \omega Y + i \sin \omega Y)}$$

and a stochastic process defined by $\boldsymbol{X(t) = \cos(\lambda t + Y)}$, Show that $\boldsymbol{\{X(t)\}}$ is stationary in the wide sense if $\boldsymbol{\emptyset(1) = \emptyset(2) = 0}$

Solution: We have $E\{X(t)\} = E\{\cos(\lambda t + Y)\} = E\{cos\lambda t\ cosY - sin\lambda t\ sinY\}$

$$= \cos\lambda t\, E(\cos Y) - \sin\lambda t\, E(\sin Y) \qquad \ldots (1)$$

Given $\emptyset(1) = 0\ i.e.,\ E\{\cos Y + i \sin Y\} = 0$

Therefore, $E(\cos Y) = 0 = E(\sin Y)$ ... (2)

Using (2) in (1), we get $E\{X(t)\} = 0$ ... (3)

$E\{X(t_1)X(t_2)\} = E\{\cos(\lambda t_1 + Y)\cos(\lambda t_2 + Y)\}$

$= \cos\lambda\, t_1 \cos\lambda\, t_2 E(cos^2 Y) + \sin\lambda\, t_1 \sin\lambda\, t_2 E(sin^2 Y) \; - \sin\lambda\,(t_1 + t_2) E(\sin Y \cos Y)$

$$= \cos\lambda\, t_1 \cos\lambda\, t_2 E\left(\frac{1}{2} + \frac{1}{2}\cos 2\,Y\right) + \sin\lambda\, t_1 \sin\lambda\, t_2\; E\left(\frac{1}{2} - \frac{1}{2}\cos 2\,Y\right)$$

$$-\frac{1}{2}\sin\lambda\,(t_1 + t_2)E(\sin 2\,Y) \qquad \dots (4)$$

Given: $\emptyset(2) = 0$, $i.e.$, $E\{\cos 2\,Y + i\sin 2\,Y\} = 0$

Therefore, $E(\cos 2\,Y) = 0 = E(\sin 2\,Y)$ ... (5)

Using (5) in (4), we get

$R(t_1, t_2) = E\{X(t_1)X(t_2)\} = \frac{1}{2}\{\cos\lambda\, t_1 \cos\lambda\, t_2 + \sin\lambda\, t_1 \sin\lambda\, t_2\}$

$$= \frac{1}{2}\cos\lambda\,(t_1 - t_2)$$

From (3) and (6), it follows that $\{X(t)\}$ is a WSS process.

Example 4: In the fair coin experiment, we define the process $\{X(t)\}$ as follows.

$X(t) = \sin \pi\, t$, if head shows, and $2t$, if tail shows.

(a) Find $E\{X(t)\}$ and (b) find $F(x, t)$ for $t = 0.25$

Solution:

(a). The probability distribution of $\{X(t)\}$ is given by

$P\{X(t) = \sin \pi t\} = \frac{1}{2}$ and $P\{X(t) = 2t\} = \frac{1}{2}$

Therefore, $E\{X(t)\} = \frac{1}{2}\sin \pi t + \frac{1}{2}.2t = \frac{1}{2}sin\,\pi t + t$

(b) When $t = 0.25$, $P\left\{X(t) = \frac{1}{\sqrt{2}}\right\} = \frac{1}{2}$ and $P\left\{X(t) = \frac{1}{2}\right\} = \frac{1}{2}$

$\therefore F(x, 0.25)$ is given by

$$F(x, 0.25) = \begin{cases} 0 & , \text{if } x < \dfrac{1}{2} \\ \dfrac{1}{2} & , \text{if } \dfrac{1}{2} \leq x < \dfrac{1}{\sqrt{2}} \\ 1 & , \text{if } \dfrac{1}{\sqrt{2}} \leq x \end{cases}$$

Example 5: If $\{X(t)\}$ is a wide – sense stationary process with autocorrelation $R(\tau) = Ae^{-\alpha|\tau|}$, determine the second order moment of the r.v $X(8) - X(5)$

Solution: Second moment of $X(8) - X(5)$ is given by

$$E[\{X(8) - X(5)\}^2] = E\{X^2(8)\} + E\{X^2(5)\} - 2E\{X(8)X(5)\} \quad \ldots (1)$$

Given: $R(\tau) = Ae^{-\alpha|\tau|}$

i.e., $R(t_1, t_2) = Ae^{-\alpha|t_1 - t_2|}$

$\therefore \quad E\{X^2(t)\} = R(t, t) = A$

$$\therefore \quad E\{X^2(8)\} = E\{X^2(5)\} = A \quad \ldots (2)$$

$$\text{Also } E\{X(8)X(5)\} = R(8, 5) = Ae^{-3\alpha} \quad \ldots (3)$$

Using (2) and (3) in (1), we get

$E[\{X(8) - X(5)\}^2] = 2A(1 - e^{-3\alpha})$

Example 6: Show that the process $X(t) = A\cos\lambda t + B\sin\lambda t$ (where A and B are r.vs) is wide – sense stationary, if

(i) $E(A) = E(B) = 0$
(ii) $E(A^2) = E(B^2)$ **and** $E(AB) = 0$

Solution: $E\{X(t)\} = \cos\lambda t E(A) + \sin\lambda t \times E(B) \quad \ldots (1)$

If $\{X(t)\}$ is to be a WSS process, $E\{X(t)\}$ must be a constant ($i.e.$, independent of t)

In (1), if $E(A)$ and $E(B)$ are any constants other than zero, $E\{X(t)\}$ will be a function of t.Therefore,

$$E(A) = E(B) = 0$$

$$R(t_1, t_2) = E\{X(t_1)X(t_2)\}$$

$$= E\{(A\cos\lambda\, t_1 + B\sin\lambda t_2)(A\cos\lambda\, t_2 + B\sin\lambda\, t_2)\}$$

$$= E(A^2)\cos\lambda\, t_1 \cos\lambda\, t_2 + E(B^2)\sin\lambda\, t_2 \sin\lambda\, t_2 +$$

$$E(AB)\sin\lambda\,(t_1 + t_2) \qquad \ldots(2)$$

If $\{X(t)\}$ is to be a WSS process, $R(t_1, t_2)$ must be a function of $(t_1 - t_2)$.

Therefore, In (2),if $E(AB) = 0$ and $E(A^2) = E(B^2) = k$, then,

$$R(t_1, t_2) = k\cos\lambda(t_1 - t_2)$$

Example 7: If the $2n$ r.vs A_r and B_r are uncorrelated with zero mean and $E(A_r^2) = E(B_r^2) = \sigma_r^2$,show that the process $X(t) = \sum_{r=1}^{n}(A_r\cos\omega_r t + B_r\sin\omega_r t)$ is wide – sense stationary. What are the mean and autocorrelation of $X(t)$?

Solution: The mean of $X(t)$ $= E\{X(t)\} = \sum_{r=1}^{n} E(A_r)\cos\omega_r t + E(B_r)\sin\omega_r t = 0$

$$E\{X(t_1)X(t_2)\} = E\left\{\sum_{r=1}^{n}\sum_{s=1}^{n}(A_r\cos\omega_r t_1 + B_r\sin\omega_r t_1)(A_s\cos\omega_s t_2 + B_s\sin\omega_s t_2)\right\}$$

Since $E\{A_r\, A_s\}, E\{B_r B_s\}, E\{A_r B_r\}$ and $E\{A_s B_s\}$ are all zero, for $r \neq s$, we have

$$R(t_1, t_2) = E\{X(t_1)X(t_2)\} = \sum_{r=1}^{n} E(A_r^2)\cos\omega_r t_1\cos\omega_r t_2 + E(B_r^2)\sin\omega_r t_1\sin\omega_r t_2$$

$$= \sum_{r=1}^{n}\sigma_r^2\cos\omega_r(t_1 - t_2)$$

Therefore, $\{X(t)\}$ is a WSS process.

5.2

Auto correlation, Cross correlation Functions and Ergodicity

Autocorrelation Function and its Properties

Definition: If the process $\{X(t)\}$ is stationary either in the *strict sense* or in the *wide sense*, then $E\{X(t)X(t-\tau)\}$ is a function of τ, denoted by $R_{xx}(\tau)$ or $R_x(\tau)$ or $R(\tau)$. This function $R(\tau)$ is called the **autocorrelation function** of the process $\{X(t)\}$.

Properties of $\boldsymbol{R(\tau)}$

1. $\boldsymbol{R(\tau)}$ is an even function of $\boldsymbol{\tau}$

Proof: We have $R(\tau) = E\{X(t)X(t-\tau)\}$

$$R(-\tau) = E\{X(t)X(t+\tau)\} = E\{X(t+\tau)X(t)\} = R(\tau)$$

Therefore, $R(\tau)$ is an even function of τ.

2. $\boldsymbol{R(\tau)}$ is maximum at $\boldsymbol{\tau = 0}$ $i.e.$, $|R(\tau)| \leq R(0)$

Proof: By Cauchy-Schwarz inequality, we have

$$\{E(XY)\}^2 \leq E(X^2)E(Y^2)$$

Put $X = X(t)$ and $Y = X(t-\tau)$

Then $[E\{X(t)X(t-\tau)\}]^2 \leq E\{X^2(t)\}E\{X^2(t-\tau)\}$

$i.e.$, $\{R(\tau)\}^2 \leq [R(0)]^2$

Taking square-root on both sides

$$|R(\tau)| \leq R(0) \text{ [since } R(0) = E\{X^2(t)\} \text{ is positive]}$$

3. If the autocorrelation function $R(\tau)$ of a real stationary process $\{X(t)\}$ is continuous at $\tau = 0$, then it is continuous at every other point.

Proof: $E[\{X(t) - X(t-\tau)\}^2]$

$$= E\{X^2(t)\} + E\{X^2(t-\tau)\} - 2E\{X(t).X(t-\tau)\}$$

$$= R(0) + R(0) - 2R(\tau) = 2[R(0) - R(\tau)] \qquad \text{...(1)}$$

Therefore, $E[\{X(t) - X(t-\tau)\}^2] = 2[R(0) - R(\tau)]$

Since $R(\tau)$ is continuous at $\tau = 0$, $\lim_{\tau \to 0} R(\tau) = R(0)$

Now, $\lim_{\tau \to 0} E[\{X(t) - X(t-\tau)\}^2] = \lim_{\tau \to 0} 2[R(0) - R(\tau)] = 0$

Thus, $\lim_{\tau \to 0} \{X(t) - X(t-\tau)\} = 0$

Therefore, $\lim_{\tau \to 0} \{X(t-\tau)\} = X(t)$

$i.e.$, $X(t)$ is continuous for all t ...(2)

Consider $R(\tau + h) - R(\tau)$

$$= E[X(t).X\{t - (\tau + h)\}] - E[X(t).X(t-\tau)]$$

$$= E\big[X(t)\big\{X\{t - \tau - h\}\big\} - X(t-\tau)\big] \qquad \text{... (3)}$$

Now, $\lim_{h \to 0} [X\{(t-\tau) - h\} - X\{(t-\tau)\}] = 0$, by (2)

$\therefore \lim_{h \to 0} \{R.S. of\ (3)\} = 0$

$\therefore \lim_{h \to 0} \{L.S. of\ (3)\} = 0$

$i.e.$, $\lim_{h \to 0} \{R(\tau + h)\} = R(\tau)$, $i.e., R(\tau)$ is continuous for all τ.

4. If $R(\tau)$ is the autocorrelation function of a stationary process $\{X(t)\}$ with no periodic component, then $\lim_{\tau \to \infty} R(\tau) = \mu_x^2$, provided the limit exists.

Proof: We have, $R(\tau) = E\{X(t)X(t-\tau)\}$.

When τ is very large, $X(t)$ and $X(t-\tau)$ are two sample functions (members) of the process $\{X(t)\}$ observed at a very long interval of time.

Therefore, $X(t)$ and $X(t-\tau)$ tend to become independent [$X(t)$ and $X(t-\tau)$ may be dependent, when $X(t)$ contains a periodic component, which is not true].

$\therefore \lim_{\tau \to \infty} R(\tau) = E\{X(t)\}.E\{X(t-\tau)\} = \mu_x^2$ [since $E\{X(t)\}$ is a constant]

$i.e., \mu_x = \sqrt{\lim_{\tau \to \infty} R(\tau)}$

Cross-Correlation Function and its Properties

Definition: If the process $\{X(t)\}$ and $\{Y(t)\}$ are jointly wide-sense stationary, then $E\{X(t)Y(t-\tau)\}$ is a function of τ, denoted by $R_{xy}(\tau)$. This function $R_{xy}(\tau)$ is called the **cross-correlation function** of the processes $\{X(t)\}$ and $\{Y(t)\}$.

We give below the properties of $R_{xy}(\tau)$ without proof. Proofs of these properties are left as exercises to the reader.

Properties

1. $R_{yx}(\tau) = R_{xy}(-\tau)$
2. $|R_{xy}(\tau)| \le \sqrt{R_{xx}(0)R_{yy}(0)}$
   This means that the maximum of $R_{xy}(\tau)$ can occur anywhere, but it cannot exceed $\sqrt{R_{xx}(0) \times R_{yy}(0)}$.
3. $|R_{xy}(\tau)| \le \frac{1}{2}\{R_{xx}(0) + R_{yy}(0)\}$
4. If the processes $\{X(t)\}$ and $\{Y(t)\}$ are orthogonal, then $R_{xy}(\tau) = 0$
5. If the processes $\{X(t)\}$ and $\{Y(t)\}$ are independent, then $R_{xy}(\tau) = \mu_x\mu_y$

Ergodicity

When we wish to take a measurement of a variable quantity in the laboratory, we usually obtain multiple measurements of the variable and average them to reduce measurement errors. If the value of the variable being measured is constant and errors are due to disturbances (noise) or due to the instability of the measuring instrument, then averaging is, in fact, a valid and useful technique. *Time averaging* is an extension of this concept, which is used in the estimation of various statistics of stochastic processes.

We normally use ensemble averages (or statistical averages) such as the *mean* and ***autocorrelation function*** for characterizing stochastic processes. To estimate ensemble averages, one has to compute a weighted average over all the member functions of the stochastic process.

For example, the ensemble mean of a discrete stochastic process $\{X(t)\}$ is computed by the formula $\mu_x = \sum x_i p_i$. If we have access only to a single sample function of the process, then we use its time-average to estimate the ensemble averages of the process.

Definition: If $\{X(t)\}$ is a stochastic process, then $\frac{1}{2T}\int_{-T}^{T} X(t)\,dt$ is called the **time-average** of $\{X(t)\}$ over $(-T, T)$ and denoted by $\overline{\boldsymbol{X_T}}$.

In general, *ensemble averages* and *time averages* are not equal except for a very special class of stochastic processes called **Ergodic processes**. *The concept of ergodicity deals with the equality of time averages and ensemble averages.*

Definition: A stochastic process $\{X(t)\}$ is said to be ***ergodic***, if its ensemble averages are equal to appropriate time averages.

This definition implies that, with probability 1, any ensemble average of $\{X(t)\}$ can be determined from a single sample function of $\{X(t)\}$.

Note: Ergodicity is a stronger condition than stationarity and hence *all stochastic processes that are stationary are not ergodic*. Moreover, ergodicity is usually defined with respect to one or more ensemble averages (such as mean and autocorrelation function) as discussed below and a process may be ergodic with respect to one ensemble average but not others.

Mean-Ergodic Process: If the stochastic process $\{X(t)\}$ has a constant mean $E\{X(t)\} = \mu$ and $\overline{X_T} = \frac{1}{2T}\int_{-T}^{T} X(t)\,dt \to \mu \text{ as } T \to \infty$, then $\{X(t)\}$ is said to be **mean-ergodic**.

Mean-Ergodic Theorem

If $\{X(t)\}$ is a stochastic process with constant mean μ and if $\overline{X_T} = \frac{1}{2T}\int_{-T}^{T} X(t)\,dt$, then $\{X(t)\}$ is mean-ergodic (or ergodic in the mean), provided

$$\lim_{T \to \infty} \{Var\overline{X_T}\} = 0.$$

Proof: $\overline{X_T} = \frac{1}{2T}\int_{-T}^{T} X(t)\,dt$

$$\therefore E\left(\overline{X_T}\right) = \frac{1}{2T}\int_{-T}^{T} E\{X(t)\}\,dt = \mu \qquad \ldots (1)$$

By Tchebycheff's inequality

$$P\{|\overline{X_T} - E(\overline{X_T})| \le \epsilon\} \ge 1 - \frac{Var(\overline{X_T})}{\epsilon^2} \qquad \ldots (2)$$

Taking limits as $T \longrightarrow \infty$ and using (1) we get

$$P\left\{\left|\lim_{T \to \infty}\{\overline{X_T}\} - \mu\right| \le \epsilon\right\} \ge 1 - \frac{\lim_{T \to \infty} Var\overline{X_T}}{\epsilon^2}$$

Therefore, When $\lim_{T \to \infty} Var\overline{X_T} = 0$, (2) becomes

$$P\left\{\left|\lim_{T \to \infty}(\overline{X_T}) - \mu\right| \le \epsilon\right\} \ge 1$$

$i.e.,$ $\lim_{T \to \infty}(\overline{X_T}) = E\{X(t)\}$ with probability 1.

Note: This theorem provides a sufficient condition for the mean-ergodicity of a stochastic process. That is, to prove the mean-ergodicity of $\{X(t)\}$, it is enough to prove $\lim_{T \longrightarrow \infty} Var(\overline{X_T}) = 0$.

Correlation Ergodic Process

The stationary process $\{X(t)\}$ is said to be **correlation ergodic** (or ergodic in the correlation), if the process $\{Y(t)\}$ is mean-ergodic, where $Y(t) = X(t+\tau)X(t)$.

That is, the stationary process $\{X(t)\}$ is correlation ergodic, if

$$\overline{Y_T} = \frac{1}{2T}\int_{-T}^{T} X(t+\tau)X(t)\,dt \text{ tends to } E\{X(t+\tau)X(t)\} = R(\tau) \text{ as } T \longrightarrow \infty.$$

Distribution Ergodic Process

If $\{X(t)\}$ is a stationary process and if $\{Y(t)\}$ is another process such that

$$Y(t) = \begin{cases} 1 & if \quad X(t) \le x \\ 0 & if \quad X(t) > x \end{cases}$$

then $\{X(t)\}$ is said to be **distribution-ergodic**, if $\{Y(t)\}$ is mean-ergodic. That is, the stationary process $\{X(t)\}$ is distribution ergodic, if

$$\overline{Y_T} = \frac{1}{2T}\int_{-T}^{T} Y(t)\,dt \to E\{Y(t)\} \text{ as } T \to \infty.$$

We note that

$$E\{Y(t)\} = 1 \times P\{X(t) \le x\} + 0 \times P\{X(t) > x\} = F_X(x)$$

Thus the stationary process $\{X(t)\}$ is distribution-ergodic, if

$$\frac{1}{2T}\int_{-T}^{T} Y(t)\,dt \to F_X(x) \text{ as } T \to \infty.$$

Example1: Given that the autocorrelation function for a stationary ergodic process with no periodic component is

$$\boldsymbol{R_{xx}(\tau) = 25 + \frac{4}{1 + 6\tau^2}}$$

Find the mean value and variance of the process $\{\boldsymbol{X(t)}\}$.

Solution: Here $R_{xx}(\tau) = R(\tau)$. By the property of autocorrelation function,

$$\mu_x^2 = \lim_{\tau \longrightarrow \infty} R_{xx}(\tau) = 25 \quad \text{(Property 4 of } R(\tau)\text{)}$$

Therefore, $\mu_x = 5$

$$E\{X^2(t)\} = R_{xx}(0) = 25 + 4 = 29$$

Therefore, $Var\{X(t)\} = E\{X^2(t)\} - (E\{X(t)\})^2 = 29 - 25 = 4$

Example2: Express the autocorrelation function of the process $\{\boldsymbol{X'(t)}\}$ in terms of the autocorrelation function of the process $\{\boldsymbol{X(t)}\}$.

Solution: Consider

$$\begin{aligned} R_{xx'}(t_1, t_2) &= \text{cross correlation function of } X(t_1) \text{ and } X'(t_2) \\ &= E\{X(t_1)X'(t_2)\} \\ &= E\left[X(t_1)\left\{\frac{X(t_2+h)-X(t_2)}{h}\right\}\right] \text{ as } h \longrightarrow 0 \\ &= \lim_{h \longrightarrow 0}\left[\frac{R_{xx}(t_1,t_2+h)-R_{xx}(t_1,t_2)}{h}\right] \\ &= \frac{\partial}{\partial t_2} R_{xx}(t_1, t_2) \qquad \dots(1) \end{aligned}$$

Similarly, $R_{x'x'}(t_1, t_2) = \frac{\partial}{\partial t_1} R_{xx'}(t_1, t_2)$ ... (2)

Using (1) in (2),

$$R_{x'x'}(t_1, t_2) = \frac{\partial^2}{\partial t_1 \partial t_2} R_{xx}(t_1, t_2) \qquad \dots.(3)$$

If $\{X(t)\}$ is a stationary process, we put $t_1 - t_2 = \tau$. From (1), (2) and (3), then we get

$$R_{xx'}(\tau) = -\frac{\partial}{\partial\tau} R_{xx}(\tau)$$

$$R_{x'x'}(\tau) = \frac{\partial}{\partial\tau} R_{xx'}(\tau) \text{ and}$$

$$R_{x'x'}(\tau) = -\frac{\partial^2}{\partial\tau^2} R_{xx}(\tau)$$

Example3: Prove that the stochastic process $\{X(t)\}$ with constant mean is mean-ergodic, if $\lim_{T\to\infty}\left[\frac{1}{4T^2}\int_{-T}^{T}\int_{-T}^{T} C(t_1,t_2)\,dt_1 dt_2\right] = 0$.

Solution: By the mean-ergodic theorem, the condition for the mean-ergodicity of the process $\{X(t)\}$ is

$\lim_{T\to\infty}\{Var(\overline{X_T})\} = 0$, where

$$\overline{X_T} = \frac{1}{2T}\int_{-T}^{T} X(t)\,dt \text{ and } E\left(\overline{X_T}\right) = E\{X(t)\} \qquad \text{(Since the mean is constant)}$$

Now $\overline{X_T}^2 = \frac{1}{4T^2}\int_{-T}^{T}\int_{-T}^{T} X(t_1)X(t_2)\,dt_1 dt_2$

$$\therefore E\left\{\overline{X_T}^2\right\} = \frac{1}{4T^2}\int_{-T}^{T}\int_{-T}^{T} R(t_1,t_2)\,dt_1 dt_2$$

$$\therefore Var\left(\overline{X_T}\right) = E\left\{\overline{X_T}^2\right\} - \left(E\left(\overline{X_T}\right)\right)^2$$

$$= \frac{1}{4T^2}\int_{-T}^{T}\int_{-T}^{T}\left[R(t_1,t_2) - E\{X(t_1)\}E\{X(t_2)\}\right]dt_1 dt_2$$

$$= \frac{1}{4T^2}\int_{-T}^{T}\int_{-T}^{T} C(t_1,t_2)\,dt_1 dt_2 \qquad \text{...(1)}$$

Therefore, the condition $\lim_{T\to\infty}\{Var(\overline{X_T})\} = 0$ is equivalent to the condition

$$\lim_{T\to\infty}\left[\frac{1}{4T^2}\int_{-T}^{T}\int_{-T}^{T}C(t_1,t_2)\,dt_1dt_2\right]=0$$

Hence the result .

Example 4: If $\overline{X_T}$ is the time – average of a stationary stochastic process $\{X(t)\}$ over $(-T,T)$, prove that $Var\left(\overline{X}_T\right)=\frac{1}{T}\int_0^{2T}C(\tau)\left[1-\frac{|\tau|}{2T}\right]d\tau$ and hence prove that the sufficient condition for the mean – ergodicity of the process $\{X(t)\}$ is

(i) $\lim_{T\to\infty}\frac{1}{T}\int_0^{2T}C(\tau)\left[1-\frac{|\tau|}{2T}\right]d\tau=0$ **and**

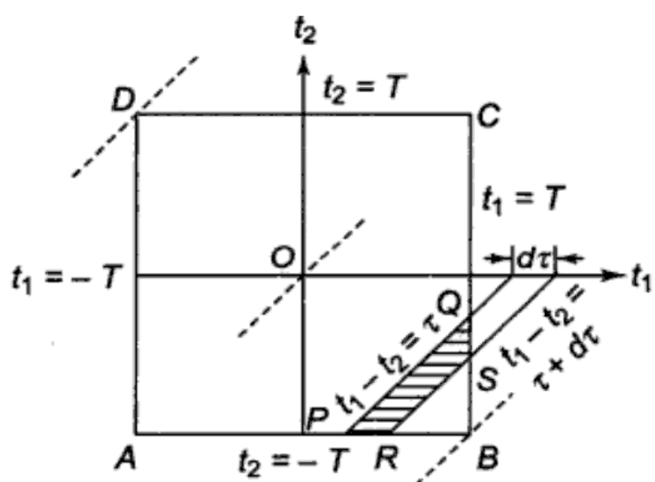
(ii) $\int_{-\infty}^{\infty}|C(\tau)|\,d\tau<\infty$

Solution: Step (1) of the Example 3 gives

$$Var\left(\overline{X}_T\right)=\frac{1}{4T^2}\int_{-T}^{T}\int_{-T}^{T}C(t_1,t_2)\,dt_1dt_2 \qquad ...(1)$$

We shall convert the double integral (1) into a single definite integral with respect to the variable $\tau=t_1-t_2$ as explained below:

The double integral (1) is evaluated over the area of the square bounded by $t_1=\pm T$ and $t_2=\pm T$ as shown in the figure.



We divide the area of the square $ABCD$ into a number of strips parallel to the line $t_1 - t_2 = 0$. Let a typical strip be $PQRS$, where PQ is given by $t_1 - t_2 = \tau$ and RS is given by $t_1 - t_2 = \tau + d\tau$.

When $PQRS$ is at the initial position D, $t_1 - t_2 = -2T$, $i.e.$, the initial value of $\tau = -2T$.

When $PQRS$ is at finial position , $t_1 - t_2 = 2T$, $i.e.$, final value of $\tau = 2T$. Hence to cover the given area $ABCD$, τ has to vary from $-2T$ to $2T$. Since $d\tau$ is very small, $C(t_1 - t_2) = C(\tau)$ can be assumed to be a constant in the strip $PQRS$.

Now $dt_1\, dt_2 =$ element area in the $t_1 t_2$ - plane

$$= \text{area of the small strip } PQRS \qquad \dots (2)$$

t_1 co-ordinate of P is obtained by solving the equations $t_1 - t_2 = \tau$ and $t_2 = -T$.

Thus $(t_1)_P = \tau - T$.

$\therefore PB(= BQ) = T - (\tau - T) = 2T - \tau$ if $\tau > 0$

$= 2T + \tau$ if $\tau < 0$

When $\tau > 0$,

Area of $PQRS =$ Area of $\Delta\, PBQ -$ Area of $\Delta\, RSB$

$$= \frac{1}{2}(2T - \tau)^2 - \frac{1}{2}(2T - \tau - d\tau)^2$$

$$= (2T - \tau)d\tau, \text{ omitting } (d\tau)^2 \qquad \dots (3)$$

From (2) and (3),

$$dt_1 dt_2 = \{2T - |\tau|\}d\tau \qquad \dots (4)$$

Using (4) in (1),

$$Var\left(\bar{X}_T\right) = \frac{1}{2T}\int_{-2T}^{2T} C(\tau)\left\{1 - \frac{|\tau|}{2T}\right\}d\tau$$

i.e., $Var\left(\bar{X}_T\right)=\frac{1}{T}\int_0^{2T} C(\tau)\left\{1-\frac{|\tau|}{2T}\right\}d\tau$ (since the integral is even) ....(5)

(i) The sufficient condition for mean – ergodicity of a stationary process $\{X(t)\}$ can also be stated as

$$\lim_{T\to\infty}\left[\frac{1}{T}\int_0^{2T} C(\tau)\left\{1-\frac{|\tau|}{2T}\right\}d\tau\right]=0$$

(ii) The sufficient condition for mean ergodicity of $\{X(t)\}$ can also given as

$$\lim_{T\to\infty}\left[\frac{1}{2T}\int_{-2T}^{2T} C(\tau)\left\{1-\frac{|\tau|}{2T}\right\}d\tau\right]=0 \quad \text{... (6)}$$

Since τ varies from $-2T$ to $2T$, $|\tau|\le 2T$.

$\therefore \quad 1-\frac{|\tau|}{2T}\le 1$

Thus, $\frac{1}{2T}\int_{-2T}^{2T} C(\tau)\left\{1-\frac{|\tau|}{2T}\right\}d\tau \le \frac{1}{2T}\int_{-2T}^{2T}|C(\tau)|d\tau$

(6) will be true, only if $\lim_{T\to\infty}\frac{1}{2T}\int_{-2T}^{2T}|C(\tau)|d\tau=0$

$i.e., if \int_{-\infty}^{\infty}|C(\tau)|d\tau$ is finite.

$i.e., if \int_{-\infty}^{\infty}|C(\tau)|d\tau<\infty$

Therefore, a sufficient condition for mean ergodicity of the stationary process $\{X(t)\}$ can also be stated as $\int_{-\infty}^{\infty}|C(\tau)|d\tau<\infty$

5.3

Power Spectral Density Function

So far we have been able to characterize a stochastic process by its mean, autocorrelation function, and covariance function. All these functions deal with time domain. We have not studied anything about the **spectral** (or **frequency domain**) properties of the process. For a deterministic signal $y(t)$, it is well known that its spectral properties are contained in its **Fourier transform** $Y(\omega)$, which is given by

$$\boldsymbol{Y}(\omega) = \int_{-\infty}^{\infty} \boldsymbol{y}(\boldsymbol{t}) \boldsymbol{e}^{-i\omega \boldsymbol{t}} \boldsymbol{dt}$$

Conversely, given $Y(\omega)$ we can recover $y(t)$ by means of the inverse Fourier transform

$$\boldsymbol{y}(\boldsymbol{t}) = \frac{\mathbf{1}}{\mathbf{2}\boldsymbol{\pi}} \int_{-\infty}^{\infty} \boldsymbol{Y}(\omega) \boldsymbol{e}^{i\omega \boldsymbol{t}} \boldsymbol{dt}$$

Thus, $Y(\omega)$ provides a complete description of $y(t)$ and vice -versa. Unfortunately, the same argument cannot be applied to a stochastic process $X(t)$ because the Fourier transform may not exist for most sample functions of the process. One of the conditions for the function $y(t)$ to be Fourier transformable is that is must be absolutely integrable,$i.e.$,

$$\int_{-\infty}^{\infty} |y(t)| dt < \infty$$

Recall that for stationary process the autocorrelation function $R(\tau)$ is bounded $i.e.$, $|R(\tau)| \leq R(0) = E[X^2(t)]$ (see property 2 of $\mathrm{R}(\tau)$ in module 5.2). Thus, instead of working directly with stochastic process $X(t)$, we work with its autocorrelation function which is bounded and hence absolutely integrable. We shall now give mathematical definition of power spectral density function of a stationary process.

Power spectral density function: If $\{X(t)\}$ is a stationary process (either in the strict sense or wide sense) with autocorrelation function $R(\tau)$, then the Fourier transform of $R(\tau)$ is called the **power spectral density function** of $\{X(t)\}$ and it is denoted by $\boldsymbol{S_{xx}(\omega)}$ or $\boldsymbol{S_x(\omega)}$ or $\boldsymbol{S(\omega)}$.

Thus,
$$\boldsymbol{S(\omega) = \int_{-\infty}^{\infty} R(\tau) e^{-i\omega\tau} d\tau} \qquad \text{... (1)}$$

Sometimes ω is replaced by $2\pi f$, where f is the frequency variable, in which case the power spectral density function will be a function of f, denoted by $S(f)$.

Then
$$\boldsymbol{S(f) = \int_{-\infty}^{\infty} R(\tau) e^{-i2\pi f\tau} d\tau} \qquad \text{... (2)}$$

Note: Equation (1) or (2) is sometimes called the **Wiener Khinchine relation**. We shall mostly follow the definition (1) and denote the power spectral density as a function of ω only.

Given the power spectral density function $S(\omega)$, the autocorrelation function $R(\tau)$ is given by the Fourier inverse transform of $S(\omega)$.

$i.e.$,
$$R(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S(\omega) e^{i\tau\omega} d\omega \qquad \text{... (3)}$$

(or)
$$R(\tau) = \int_{-\infty}^{\infty} S(f) e^{i2\pi\tau f} df \qquad \text{... (4)}$$

If $\{X(t)\}$ and $\{Y(t)\}$ are two jointly stationary random processes with cross-correlation function $R_{xy}(\tau)$, then the Fourier transform of $R_{xy}(\tau)$ is called the **crosspower spectral density** of $\{X(t)\}$ and $\{Y(t)\}$, denoted as $\boldsymbol{S_{xy}(\omega)}$.

$i.e.$,
$$S_{xy}(\omega) = \int_{-\infty}^{\infty} R_{xy}(\tau) e^{-i\omega\tau} d\tau$$

Properties of Power Spectral Density Function:

1. The value of the spectral density function at zero frequency is equal to the total area under the graph of the autocorrelation function. By putting $\omega = 0$ in (1) or $f = 0$ in (2), we get
$$S(0)=\int_{-\infty}^{\infty} R(\tau)d\tau$$, which is the given property.
2. The mean square value of a wide-sense stationary process is equal to the total area under the graph of the spectral density. By putting $\tau = 0$ in (4), we get
$$E\left[X^2(t)\right]=R(0)=\int_{-\infty}^{\infty} S(f)df$$, which is the given property
3. The spectral density function of a real stochastic process is an even function **(For proof see P1).**
4. The spectral density of a process $\{X(t)\}$ is a real function of ω and non-negative **(For proof see P2).**
5. Spectral density of any WSS is non-negative $i.e.$, $S(\omega) \geq 0$ **(see example 9).**
6. The spectral density and autocorrelation function of a real WSS process form a Fourier cosine transform pair**(For proof see P3)**

Wiener-Khinchine Theorem

If $X_T(\omega)$ is the Fourier transform of the truncated stochastic process defined as

$$X_T(t) = \begin{cases} X(t) & for \quad |t| \leq T \\ 0 & for \quad |t| > T \end{cases}$$

Where $\{X(t)\}$ is a real WSS process with power spectral density function $S(\omega)$, then

$$S(\omega) = lim_{T\to\infty}\left[\frac{1}{2T}E\{|X_T(\omega)|^2\}\right]$$

Proof: **See P4**

Example 1: The autocorrelation function of the random telegraph signal process is given by $R(\tau) = a^2 e^{-2\gamma|\tau|}$. Determine the power density spectrum of the random telegraph signal.

Solution: $S(\omega) = \int_{-\infty}^{\infty} R(\tau) e^{-i\omega\tau} d\tau$

$$= a^2 \int_{-\infty}^{\infty} e^{-2\gamma|\tau|} (\cos\omega\tau - i\sin\omega\tau) d\tau$$

$$= 2a^2 \int_0^{\infty} e^{-2\gamma\tau} \cos\omega\tau d\tau$$

$$= \left[\frac{2a^2 e^{-2\gamma\tau}}{4\gamma^2 + \omega^2} (-2\gamma\cos\omega\tau + \omega\sin\omega\tau) \right]_0^{\infty}$$

$$= \frac{4a^2\gamma}{4\gamma^2 + \omega^2}$$

Example 2: The autocorrelation function of the Poisson increment process is given by

$$R(\tau) = \begin{cases} \lambda^2 & for \quad |\tau| > \epsilon \\ \lambda^2 + \frac{\lambda}{\epsilon}\left(1 - \frac{|\tau|}{\epsilon}\right) & for \quad |\tau| \le \epsilon \end{cases}$$

Prove that its spectral density function is given by

$$S(\omega) = 2\pi\lambda^2\delta(\omega) + \frac{4\lambda \sin^2\left(\frac{\omega\epsilon}{2}\right)}{\epsilon^2\omega^2}$$

Solution:

$$S(\omega)=\int_{-\varepsilon}^{\varepsilon}\left\{\lambda^2+\frac{\lambda}{\varepsilon}\left(1-\frac{|\tau|}{\varepsilon}\right)\right\}e^{-i\omega\tau}d\tau+\int_{-\infty}^{-\varepsilon}\lambda^2e^{-i\omega\tau}d\tau+\int_{\varepsilon}^{\infty}\lambda^2e^{-i\omega\tau}d\tau$$

$$=\frac{\lambda}{\varepsilon}\int_{-\varepsilon}^{\varepsilon}\left(1-\frac{|\tau|}{\varepsilon}\right)e^{-i\omega\tau}d\tau+\int_{-\infty}^{-\varepsilon}\lambda^2e^{-i\omega\tau}d\tau+\int_{-\varepsilon}^{\varepsilon}\lambda^2e^{-i\omega\tau}d\tau+\int_{\varepsilon}^{\infty}\lambda^2e^{-i\omega\tau}d\tau$$

$$=\frac{\lambda}{\varepsilon}\int_{-\varepsilon}^{\varepsilon}\left(1-\frac{|\tau|}{\varepsilon}\right)e^{-i\omega\tau}d\tau+\int_{-\infty}^{\infty}\lambda^2e^{-i\omega\tau}d\tau$$

$$=\frac{2\lambda}{\varepsilon}\int_{0}^{\varepsilon}\left(1-\frac{\tau}{\varepsilon}\right)\cos\omega\tau d\tau+F\left\{\lambda^2\right\}$$

where $F(\lambda^2)$ is the Fourier transform of λ^2.

$$=\frac{2\lambda}{\varepsilon}\left[\left(1-\frac{\tau}{\varepsilon}\right)\frac{\sin\omega\tau}{\omega}+\frac{1}{\varepsilon}\left(\frac{-\cos\omega\tau}{\omega^2}\right)\right]_0^{\varepsilon}+F\left\{\lambda^2\right\} \qquad (Integeration\ by\ parts)$$

$$=\frac{2\lambda}{\varepsilon^2\omega^2}(1-\cos\omega\varepsilon)+F\left\{\lambda^2\right\}$$

$$=\frac{4\lambda\sin^2\left(\frac{\omega\varepsilon}{2}\right)}{\varepsilon^2\omega^2}+F\left\{\lambda^2\right\} \qquad .....(1)$$

The Fourier inverse transform of $S(\omega)$ is given by

$$R(\tau)=F^{-1}\left\{S(\omega)\right\}$$

$$=\frac{1}{2\pi}\int_{-\infty}^{\infty}S(\omega)e^{i\tau\omega}d\omega$$

Let us now find $R(\tau)$ corresponding to $S(\omega)=2\pi\lambda^2\delta(\omega)$, where $\delta(\omega)$ is the **unit impulse function**.

$i.e.,\qquad R(\tau)=F^{-1}\left\{2\pi\lambda^2\delta(\omega)\right\}$

$$= \frac{2\pi\lambda^2}{2\pi}\int_{-\infty}^{\infty} \delta(\omega) e^{i\tau\omega} d\omega$$

$$= \lambda^2 \left[\text{since } \int_{-\infty}^{\infty} \phi(t)\delta(t)dt = \phi(0)\right]$$

Therefore, $F(\lambda^2) = 2\pi\lambda^2\delta(\omega)$ …(2)

Inserting (2) in (1) the required result is obtained.

Example 3: Find the power spectral density function of a WSS process with autocorrelation function

$$\boldsymbol{R(\tau) = e^{-\alpha\tau^2}}$$

Solution:

$$S(\omega) = \int_{-\infty}^{\infty} e^{-\alpha\tau^2} e^{-i\omega\tau} d\tau$$

$$= e^{-\frac{\omega^2}{4\alpha}} \int_{-\infty}^{\infty} e^{-\alpha\left(\tau + \frac{i\omega}{2\alpha}\right)^2} d\tau$$

$$= \frac{1}{\sqrt{\alpha}} e^{-\frac{\omega^2}{4\alpha}} \int_{-\infty}^{\infty} e^{-x^2} dx \text{ , } putting\ \sqrt{\alpha}\left(\tau + \frac{i\omega}{2\alpha}\right) = x$$

$$= \sqrt{\frac{\pi}{\alpha}} e^{-\frac{\omega^2}{4\alpha}} \left[\text{since } \int_{-\infty}^{\infty} e^{-x^2} dx = \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}\right]$$

Example 4: A stochastic process $\{X(t)\}$ is given by $X(t) = A\cos pt + B\sin pt$, where A and B are independent r.vs such that $E(A) = E(B) = 0$ and $E(A^2) = E(B^2) = \sigma^2$. Find the power spectral density of the process.

Solution: The autocorrelation function of the given process can be found as

$$R(\tau) = \sigma^2 \cos p\tau$$

$$S(\omega) = \int_{-\infty}^{\infty} \sigma^2 \cos p\tau e^{-i\omega\tau} d\tau \qquad \text{…(1)}$$

Consider $F^{-1}\{\pi\sigma^2[\delta(\omega + p) + \delta(\omega - p)]\}$

$$= \frac{1}{2\pi} \pi\sigma^2 \int_{-\infty}^{\infty} \left[\delta(\omega + p) + \delta(\omega - p) \right] e^{i\tau\omega} d\omega$$

$$= \frac{\sigma^2}{2} \left[e^{-i\tau p} + e^{i\tau p} \right] \quad \left\{ \text{since } \int_{-\infty}^{\infty} \phi(t)\delta(t-a)\,dt = \phi(a) \right\}$$

$$= \sigma^2 \cos p\tau$$

$$\therefore F(\sigma^2 \cos p\tau) = \pi\sigma^2[\delta(\omega + p) + \delta(\omega - p)] \qquad ...(2)$$

Using (2) in (1), we get,

$$S(\omega) = \pi\sigma^2 \left[\delta(\omega + p) + \delta(\omega - p) \right]$$

Example 5: If $\boldsymbol{Y(t) = X(t+a) - X(t-a)}$, prove that $\boldsymbol{R_{yy}(\tau) = 2R_{xx}(\tau) - R_{xx}(\tau + 2a) - R_{xx}(\tau - 2a)}$. Hence prove that $\boldsymbol{S_{yy}(\omega) = 4\sin^2 a\omega\, S_{xx}(\omega)}$.

Solution: $R_{yy}(\tau) = 2R_{xx}(\tau) - R_{xx}(\tau + 2a) - R_{xx}(\tau - 2a)$

Taking Fourier transforms on both sides.

$$S_{yy}(\omega) = 2S_{xx}(\omega) - \int_{-\infty}^{\infty} R_{xx}(\tau + 2a) e^{-i\omega\tau} d\tau - \int_{-\infty}^{\infty} R_{xx}(\tau - 2a) e^{-i\omega\tau} d\tau$$

$$= 2S_{xx}(\omega) - e^{i2a\omega} \int_{-\infty}^{\infty} R_{xx}(u) e^{-i\omega u} du - e^{-i2a\omega} \int_{-\infty}^{\infty} R_{xx}(v) e^{-i\omega v} dv$$

(putting $\tau + 2a = u$ in the first integral and $\tau - 2a = v$ in the second integral)

$i.e., S_{yy}(\omega) = 2S_{xx}(\omega) - \{e^{i2a\omega} + e^{-i2a\omega}\} S_{xx}(\omega)$

$$= 2(1 - \cos 2a\omega) S_{xx}(\omega)$$

$$= 4\sin^2 a\omega\, S_{xx}(\omega)$$

Example 6: If the process $\{X(t)\}$ is defined as $X(t) = Y(t)Z(t)$, where $\{Y(t)\}$ and $\{Z(t)\}$ are independent WSS processes, prove that

(i) $\boldsymbol{R_{xx}(\tau) = R_{yy}(\tau) R_{zz}(\tau)}$ **and**

(ii) $\boldsymbol{S_{xx}(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{yy}(\alpha) S_{zz}(\omega - \alpha) d\alpha}$

Solution: $S_{xx}(\omega) = F\{R_{xx}(\tau)\} = F\{R_{yy}(\tau)R_{zz}(\tau)\}$ ....(1)

Consider $F^{-1}\left[\int_{-\infty}^{\infty} S_{yy}(\alpha)S_{zz}(w-\alpha)\,d\alpha\right]$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} S_{yy}(\alpha)S_{zz}(\omega-\alpha)e^{i\omega\tau}\,d\alpha\,d\omega$$

Putting $\alpha = y$ and $\omega - \alpha = z$, we get (from calculus)

$$d\alpha d\omega = \begin{vmatrix} \alpha_y & \alpha_z \\ \omega_y & \omega_z \end{vmatrix} dy\,dz = \begin{vmatrix} 1 & 0 \\ 1 & 1 \end{vmatrix} dy\,dz$$

$$\begin{aligned}
\therefore F^{-1}&\left[\int_{-\infty}^{\infty} S_{yy}(\alpha)S_{zz}(\omega-\alpha)\,d\alpha\right] \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} S_{yy}(y)S_{zz}(z)e^{i(y+z)\tau}\,dydz \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{yy}(y)e^{iy\tau}dy\int_{-\infty}^{\infty} S_{zz}(z)e^{iz\tau}dz \\
&= F^{-1}\{S_{yy}(\omega)\}\,2\pi F^{-1}\{S_{zz}(\omega)\} \\
&= 2\pi R_{yy}(\tau)R_{zz}(\tau)
\end{aligned}$$

$$\therefore F\{R_{yy}(\tau)R_{zz}(\tau)\} = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{yy}(\alpha)S_{zz}(\omega-\alpha)\,d\alpha \quad \text{.... (2)}$$

Using (2) in (1), we get $S_{xx}(\omega)$ in the required form.

Example 7: If the power spectral density of a WSS process is given by

$$\boldsymbol{S(\omega)} = \begin{cases} \boldsymbol{\frac{b}{a}(a-|\omega|)} & , \quad \boldsymbol{|\omega| \le a} \\ \boldsymbol{0} & , \quad \boldsymbol{|\omega|} > a \end{cases}$$

Find the autocorrelation function of the process.

Solution: The autocorrelation function

$$R(\tau)=F^{-1}\{S(\omega)\}$$

$$=\frac{1}{2\pi}\int_{-\infty}^{\infty}S(\omega)e^{i\tau\omega}d\omega$$

$$=\frac{1}{2\pi}\int_{-a}^{a}\frac{b}{a}(a-|\omega|)e^{i\tau\omega}d\omega=\frac{1}{2\pi}\int_{-a}^{a}\frac{b}{a}(a-|\omega|)\cos\tau\omega\,d\omega$$

$$=\frac{1}{\pi}\int_{0}^{a}\frac{b}{a}(a-\omega)\cos\tau\omega\,d\omega$$

$$=\frac{b}{\pi a}\left\{(a-\omega)\frac{\sin\tau\omega}{\tau}-\frac{\cos\tau\omega}{\tau^2}\right\}_0^a \qquad \text{(integration by parts)}$$

$$=\frac{b}{\pi a\tau^2}(1-\cos a\tau)$$

$$=\frac{ab}{2\pi}\left(\frac{\sin a\frac{\tau}{2}}{a\frac{\tau}{2}}\right)^2$$

Example 8: The power spectral density function of a zero mean WSS process $\{X(t)\}$ is given by

$$\boldsymbol{S(\omega)=\begin{cases}1 & , \quad |\omega|<\omega_0\\ 0 & , \quad elsewhere\end{cases}}$$

Find $R(\tau)$ and show also that $X(t)$ and $X\left(t+\frac{\tau}{\omega_0}\right)$ are uncorrelated.

Solution:

We have $R(\tau)=F^{-1}\{S(\omega)\}$

$$i.e.,\ R(\tau)=\frac{1}{2\pi}\int_{-\infty}^{\infty}S(\omega)e^{i\tau\omega}d\omega=\frac{1}{2\pi}\int_{-\omega_0}^{\omega_0}e^{i\tau\omega}d\omega$$

$$=\frac{1}{2\pi}\left\{\frac{e^{i\tau\omega}}{i\tau}\right\}_{-\omega_0}^{\omega_0}=\frac{1}{2\pi i\tau}\left(e^{i\tau\omega_0}-e^{-i\tau\omega_0}\right)$$

$$=\frac{1}{\pi\tau}\sin\omega_0\tau$$

Now, $E\left\{X\left(t+\frac{\pi}{\omega_0}\right)X(t)\right\} = R\left(\frac{\pi}{\omega_0}\right) = \frac{\omega_0}{\pi^2} sin\left(\omega_0 \frac{\pi}{\omega_0}\right) = \frac{\omega_0}{\pi^2} sin\pi = 0$

Since the mean of the process is zero,

$$C\left\{X\left(t+\frac{\pi}{\omega_0}\right)X(t)\right\} = E\left\{X\left(t+\frac{\pi}{\omega_0}\right)X(t)\right\} = 0$$

Therefore, $X(t)$ and $X\left(t+\frac{\pi}{\omega_0}\right)$ are uncorrelated.

Example 9: Property (5) of power spectral density. Prove that the spectral density of any $\boldsymbol{WSS}$ process is non – negative. $\boldsymbol{i.e.}$, $\boldsymbol{S(w) \geq 0}$.

Solution: If possible, let $S(\omega) < 0$ at $\omega = \omega_0$. That is, let $S(\omega) < 0$ in $\omega_0 - \frac{\epsilon}{2} < \omega < \omega_0 + \frac{\epsilon}{2}$, where ϵ is very small. Let us assume that the system function of the convolution type linear system is

$$H(\omega) = \begin{cases} 1\,, & \omega_0 < \frac{\epsilon}{2} < \omega < \omega_0 + \frac{\epsilon}{2} \\ 0\,, & otherwise \end{cases}$$

Note: In this case, system is called a **narrow band filter**

Now $\quad S_{yy}(\omega) = |H(\omega)|^2 S_{xx}(\omega)$

$$= \begin{cases} S_{xx}(\omega)\,, & \omega_0 - \frac{\epsilon}{2} < \omega < \omega_0 + \frac{\epsilon}{2} \\ 0 \quad , & elsewhere \end{cases}$$

$E\{Y^2(t)\} = R_{yy}(0)$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{yy}(\omega)\, d\omega$$

$$= \frac{1}{2\pi}\int_{\omega_0-\frac{\varepsilon}{2}}^{\omega+\frac{\varepsilon}{2}} S_{xx}(\omega)\, d\omega$$

$$= \frac{\epsilon}{2\pi} S_{xx}(\omega_0)$$

[Since $S_{xx}(\omega_0)$ can be considered a contant $S_{xx}(\omega_0)$, as the band is narrow]

Since $E\{Y^2(t)\} \geq 0$, $S_{xx}(\omega_0) \geq 0$, which is contrary to our initial assumption. Therefore $S_{xx}(\omega) \geq 0$, since $\omega = \omega_0$ is arbitrary.

5.4

Linear Systems with Random Inputs

Mathematically, a system is a functional relationship between the input $x(t)$ and the output $y(t)$. Usually this relationship is written as $y(t) = f[x(t)]$, $-\infty < t < \infty$.

If we assume that $x(t)$ represents a sample function of a random process $\{X(t)\}$, the system produces an output or response $y(t)$ and the ensemble of the output functions forms a random process $\{Y(t)\}$. The process $\{Y(t)\}$ can be considered as the output of the system or transformation f with $\{X(t)\}$ as the input. The system is completely specified by the operator f.

We recall that $X(t)$ actually means $X(s,t)$, where $s \in S$ (sample space). If the system operates only on the variable t treating s as a parameter, it is called a ***deterministic system***. If the system operates on both t and s, it is called ***stochastic***. We shall consider only deterministic systems in our study.

Definitions: If $f[a_1X_1(t) \pm a_2X_2(t)] = a_1f[X_1(t)] \pm a_2f[X_2(t)]$, then f is called a **linear system**.

If $Y(t+h) = f[X(t+h)]$, where $Y(t) = f[X(t)]$, f is called a **time-invariant system** or **$\boldsymbol{X(t)}$ and $\boldsymbol{Y(t)}$ are said to form a *time-invariant system*.**

If the output $Y(t_1)$ at a given time $t = t_1$ depends only on $X(t_1)$ and not on any other past or future values of $X(t)$, then the system f is called a **memoryless system.**

If the value of the output $Y(t)$ at $t = t_1$ depends only on the past values of the input $X(t)$, $t \le t_1$, i.e., $Y(t_1) = f[X(t); t \le t_1]$, then the system is called a **causal system.**

System in the Form of Convolution

Very often in electrical systems, the output $Y(t)$ is expressed as a convolution of the input $X(t)$ with a system weighting function $h(t)$, i.e., the input-output relationship will be of form

$$Y(t) = \int_{-\infty}^{\infty} h(u) X(t-u) du \qquad \text{... (1)}$$

Unit Impulse Response of the System

The unit impulse function $\delta(t-a)$ is defined as

$$\delta(t-a) = \begin{cases} \frac{1}{\epsilon} & if \;\; a - \frac{\epsilon}{2} \leq t \leq a + \frac{\epsilon}{2} \\ 0 & otherwise \end{cases}$$

where $\epsilon \longrightarrow 0$.

Let $\phi(t)$ be some bounded function of t such that it can be considered as a constant in a small interval of length ϵ.

Then $\int_{-\infty}^{\infty} \phi(t)\delta(t-a)dt = \int_{a-\frac{\varepsilon}{2}}^{a+\frac{\varepsilon}{2}} \phi(t)\frac{1}{\varepsilon}dt$

$$= \frac{\phi(a)}{\varepsilon}\int_{a-\frac{\varepsilon}{2}}^{a+\frac{\varepsilon}{2}} dt = \frac{\phi(a)}{\varepsilon}\varepsilon = \phi(a)$$

Thus, $\int_{-\infty}^{\infty} \phi(t)\delta(t-a)dt = \phi(a)$.

If we take $a = 0$, we get

$$\int_{-\infty}^{\infty} \phi(t)\delta(t)dt = \phi(0) \qquad \text{... (2)}$$

Put $X(t) = \delta(t)$ in (1), then

$$\begin{aligned} Y(t) &= \int_{-\infty}^{\infty} h(u)\delta(t-u)du \\ &= \int_{-\infty}^{\infty} h(t-u)\delta(u)du \;\; \text{(by the property of the convolution)} \\ &= h(t-0) \text{, by (2)} \\ &= h(t) \end{aligned}$$

Thus if the input of the system is the unit impulse function, then the output or response is the system weighting function. Hence the system weighting function $h(t)$ will be hereafter called ***unit impulse response function***.

Properties

1. *If a system is such that its input $X(t)$ and its output $Y(t)$ are related by a convolution integral, i.e., if $Y(t)=\int_{-\infty}^{\infty} h(u)X(t-u)du$, then the system is a linear time-invariant system.*

Proof: Let $X(t) = a_1X_1(t) + a_2X_2(t)$. Then

$$\begin{aligned} Y(t) &= \int_{-\infty}^{\infty} h(u)\left[a_1X_1(t-u)+a_2X_2(t-u)\right]du \\ &= a_1Y_1(t)+a_2Y_2(t) \end{aligned}$$

Therefore, the system is linear. If $X(t)$ is replaced by $X(t+h)$, then

$$\int_{-\infty}^{\infty} h(u)X\left(\overline{t+h}-u\right)du = Y(t+h)$$

Therefore, the system is time-invariant.

Note: If $h(t)$ is absolutely integrable, viz., $\int_{-\infty}^{\infty} |h(t)|dt < \infty$, then the system is said to be *stable* in the sense that every bounded input gives a bounded output.

In addition, if $h(t) = 0$, when $t < 0$, the system is said to be **causal.**

2. *If the input to a time-invariant, stable linear system is a WSS process, then the output will also be a WSS process.* (For proof **see P1**)
3. *If $\{X(t)\}$ is a WSS process and if $Y(t)=\int_{-\infty}^{\infty} h(u)X(t-u)du$, then*
   (i) $R_{xy}(\tau) = R_{xx}(\tau) * h(-\tau)$ *and*
   (ii) $R_{yy}(\tau) = R_{xy}(\tau) * h(\tau)$, *where $*$ denotes convolution. Also*
   (iii) $S_{xy}(\omega) = S_{xx}(\omega)H^*(\omega)$ *and*
   (iv) $S_{yy}(\omega) = S_{xx}(\omega)|H(\omega)|^2$

(For proof **see P2**)

4. *If $\{X(t)\}$ is a WSS process and if $Y(t)=\int_{-\infty}^{\infty} h(u)X(t-u)du$, then*

$$R_{yy}(\tau) = R_{xx}(\tau) * K(\tau)$$

where $K(t) = h(t)h(-t) = \int_{-\infty}^{\infty} h(u)h(t+u)\,du$ (For proof **see P3**)

5. *The power spectral densities of the input and output processes in the system are connected by the relation*

$$S_{yy}(\omega) = |H(\omega)|^2 S_{xx}(\omega),$$

where $H(\omega)$ is the Fourier transform of unit impuse response function $h(t)$.

(For proof **see P4**)

Example 1: The short-time moving average of a process $\{X(t)\}$ is defined as $Y(t) = \frac{1}{T}\int_{t-T}^{t} X(s)\,ds$. Prove that $X(t)$ and $Y(t)$ are related by means of a convolution type integral. Find the unit impulse response of the system also.

Solution: We have $Y(t) = \frac{1}{T}\int_{t-T}^{t} X(s)\,ds$ ... (1)

Putting $s = t - u$ and treating t as a parameter, (1) becomes

$$Y(t) = \frac{1}{T}\int_{0}^{T} X(t-u)\,du \qquad \text{... (2)}$$

Let us define the unit impulse response of the system as follows:

$$h(t) = \begin{cases} \frac{1}{T} & , \quad for\ \ 0 \le t \le T \\ 0 & , \quad otherwise \end{cases}$$

Then (2) can be expressed as

$$Y(t) = \int_{-\infty}^{\infty} h(u) X(t-u)\,du$$

which is a convolution type integral.

Example 2: If the input $x(t)$ and the output $y(t)$ are connected by the differential equation $T\frac{dy(t)}{dt} + y(t) = x(t)$, then prove that they can be related by means of a convolution type integral. Assume that $x(t)$ and $y(t)$ are zero for $t \le 0$.

Solution: The given differential equation $y'(t) + \frac{1}{T}y(t) = \frac{1}{T}x(t)$ is a linear equation. Its solution is

$$y(t)e^{\frac{t}{T}} = \int \frac{1}{T}x(u)e^{\frac{u}{T}}du + c$$

i.e.,
$$y(t)e^{\frac{t}{T}} = \frac{1}{T}\int x(u)e^{-\frac{t-u}{T}}du + c$$

Since $y(0) = 0$,

$$y(t) = \frac{1}{T}\int_0^t x(u)e^{-\frac{t-u}{T}}du$$

(or)
$$y(t) = \frac{1}{T}\int_0^t x(t-u)e^{-\frac{u}{T}}du \qquad \text{... (1)}$$

Given:

$$x(t) = 0, \text{ for } t < 0$$

$$\therefore x(t-u) = 0, \text{ for } t < u$$

$\therefore$ (1) can be written as

$$y(t) = \frac{1}{T}\int_0^\infty x(t-u)e^{-\frac{u}{T}}du \qquad \text{... (2)}$$

Now if we define

$$h(t) = \begin{cases} \frac{1}{T}e^{-\frac{t}{T}} & , \quad for\ \ t \geq 0 \\ 0 & , \quad otherwise \end{cases}$$

(2) can be rewritten as

$$y(t) = \int_{-\infty}^{\infty} h(u)x(t-u)du$$

Hence the result.

Example 3: $\boldsymbol{X(t)}$ **is the input voltage to a circuit (system) and** $\boldsymbol{Y(t)}$ **is the output voltage.** $\boldsymbol{\{X(t)\}}$ **is a stationary random process with** $\boldsymbol{\mu_x = 0}$ **and** $\boldsymbol{R_{xx}(\tau) = e^{-\alpha|\tau|}}$**. Find** $\boldsymbol{\mu_y}$ **,** $\boldsymbol{S_{yy}(\omega)}$ **and** $\boldsymbol{R_{yy}(\tau)}$**, if the power transfer function is**

$$\boldsymbol{H(\omega) = \frac{R}{R + iL\omega}}$$

Solution: $Y(t) = \int_{-\infty}^{\infty} h(\alpha) X(t-\alpha) d\alpha$

$$\therefore E\{Y(t)\} = \int_{-\infty}^{\infty} h(\alpha) E\{X(t-\alpha)\} d\alpha = 0$$

Since $[E\{X(t-\alpha) = \mu_x = 0\}]$

$$S_{xx}(\omega) = \int_{-\infty}^{\infty} R_{xx}(\tau) e^{-i\omega\tau} d\tau$$

$$= \int_{-\infty}^{0} e^{\alpha\tau} e^{-i\omega\tau} d\tau + \int_{0}^{\infty} e^{-\alpha\tau} e^{-i\omega\tau} d\tau$$

$$= \left\{ \frac{e^{(\alpha - i\omega)\tau}}{\alpha - i\omega} \right\}_{-\infty}^{0} + \left\{ \frac{e^{-(\alpha+i\omega)\tau}}{-(\alpha+i\omega)} \right\}_{0}^{\infty}$$

$$= \frac{1}{\alpha - i\omega} + \frac{1}{\alpha + i\omega} = \frac{2\alpha}{\alpha^2 + \omega^2}$$

Now, $S_{yy}(\omega) = S_{xx}(\omega)|H(\omega)|^2$

$$= \frac{2\alpha}{\alpha^2+\omega^2} \frac{R^2}{R^2+L^2\omega^2}$$

$$= \frac{\{(2\alpha R^2/(R^2-L^2\alpha^2))\}}{\alpha^2+\omega^2} + \frac{\{(2\alpha R^2/(\alpha^2-R^2/L^2))\}}{R^2+L^2\omega^2} \quad \text{(by partial fractions)}$$

$$= \frac{2\alpha\left(\frac{R}{L}\right)^2}{\left(\frac{R}{L}\right)^2 - \alpha^2} \times \frac{1}{\alpha^2+\omega^2} + \frac{2\alpha R^2/L^2}{\alpha^2 - \left(\frac{R}{L}\right)^2} \times \frac{1}{\left(\frac{R}{L}\right)^2 + \omega^2}$$

$$= \lambda \frac{1}{\alpha^2+\omega^2} + \mu \frac{1}{\left(\frac{R}{L}\right)^2 + \omega^2}, \text{ say}$$

$$\therefore R_{yy}(\tau)=\frac{\lambda}{2\pi}\int_{-\infty}^{\infty}\frac{e^{i\omega\tau}}{\alpha^2+\omega^2}d\omega+\frac{\mu}{2\pi}\int_{-\infty}^{\infty}\frac{e^{i\omega\tau}}{\left(\frac{R}{L}\right)^2+\omega^2}d\omega \qquad \ldots (1)$$

We can prove that, by contour integration technique,

$$\int_{-\infty}^{\infty}\frac{e^{iaz}}{z^2+b^2}dz=\frac{\pi}{b}e^{-ab};\ a>0 \qquad \ldots (2)$$

Using (2) in (1)

$$R_{yy}(\tau)=\frac{\left(\frac{R}{L}\right)^2}{\left(\frac{R}{L}\right)^2-\alpha^2}e^{-\alpha|\tau|}+\frac{\left(\frac{R}{L}\right)^2\alpha}{\alpha^2-\left(\frac{R}{L}\right)^2}e^{-\left(\frac{R}{L}\right)|\tau|}$$

Example 4: Given that $Y(t)=\frac{1}{2\varepsilon}\int_{t-\varepsilon}^{t+\varepsilon}X(\alpha)d\alpha$**, where** $\{\boldsymbol{X(t)}\}$ **is a WSS process, prove that** $\boldsymbol{S_{yy}(\omega)=\frac{\sin^2\epsilon\omega}{\epsilon^2\omega^2}S_{xx}(\omega)}$ **and hence find the relation between** $\boldsymbol{R_{xx}(\tau)}$ **and** $\boldsymbol{R_{yy}(\tau)}$**.**

Solution: Putting $\alpha=t-u$, we get $Y(t)=\frac{1}{2\varepsilon}\int_{-\varepsilon}^{\varepsilon}X(t-u)du$

If we define $h(t)$ as follows

$$h(t)=\begin{cases}\frac{1}{2\epsilon} & , \quad |t|\le\epsilon\\ 0 & , \quad |t|>\epsilon\end{cases}$$

then $Y(t)=\int_{-\infty}^{\infty}h(u)X(t-u)du$

$\therefore S_{yy}(\omega)=|H(\omega)|^2S_{xx}(\omega)$, where $H(\omega)=F\{h(t)\}$

$$=\int_{-\varepsilon}^{\varepsilon}\frac{1}{2\varepsilon}e^{-i\omega t}dt=\frac{\sin\varepsilon\omega}{\varepsilon\omega}$$

i.e., $S_{yy}(\omega)=\frac{\sin^2\epsilon\omega}{(\epsilon\omega)^2}S_{xx}(\omega)$

$$\therefore R_{yy}(\tau) = F^{-1}\left\{\frac{\sin^2 \epsilon\omega}{\epsilon^2\omega^2} S_{xx}(\omega)\right\} \qquad \text{(inverse Fourier transformation)}$$

$$= F^{-1}\left\{\frac{\sin^2 \epsilon\omega}{\epsilon^2\omega^2}\right\} * R_{xx}(\tau) \qquad \dots (1)$$

We can prove that

if $$R(\tau) = \begin{cases} 1 - \frac{|\tau|}{2\epsilon} & , \quad if \;\; |\tau| \le 2\epsilon \\ 0 & , \quad if \; |\tau| > 2\epsilon \end{cases}$$

then $S(\omega) = 2\epsilon \; \frac{\sin^2 \epsilon\omega}{\epsilon^2\omega^2}$

$$\therefore F^{-1}\left(\frac{\sin^2 \epsilon\omega}{\epsilon^2\omega^2}\right) = \begin{cases} \frac{1}{2\epsilon}\left(1 - \frac{|\tau|}{2\epsilon}\right) & , \quad if \;\; |\tau| \le 2\epsilon \\ 0 & , \quad if \; |\tau| > 2\epsilon \end{cases} \qquad \dots (2)$$

Using (2) in (1)

$$R_{yy}(\tau) = \frac{1}{2\varepsilon}\int_{-2\varepsilon}^{2\varepsilon}\left(1 - \frac{|u|}{2\varepsilon}\right) R_{xx}(\tau - u)\, du$$

5.5

Random walk and Telegraph signal processes

Random walk: A random walk is derived from a sequence of Bernoulli trials as follows:

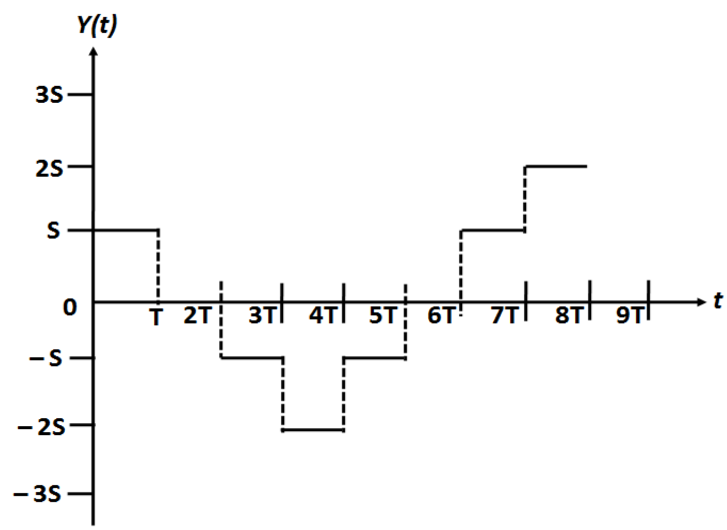
Consider a Bernoulli trial in which the probability of success is p and the probability of failure is $1-p=q$. Assume that the experiment is performed every T time units, and let the random variable X_k denote the outcome of the k^{th} trial. Furthermore, assume that the p.m.f of X_k is as follows:

$$P_{X_k}(x)=\begin{cases} p & , \ x=1 \\ 1-p & , \ x=-1 \end{cases}$$

Finally, let the random variable Y_n be defined as follows:

$$Y_n=\sum_{k=1}^{n} X_k \qquad n=1,\ 2,\ \ldots$$

where $Y_0=0$. If we use X_k to model a process where we take a step to the right if the outcome of the k^{th} trial is a success and a step to the left if the outcome is a failure, then the random variable Y_n represents the location of the process relative to the starting point (or origin) at the end of the n^{th} trial.



A sample Path of the Random Walk

The resulting trajectory of the process as it moves through the xy plane, where the x coordinate represents the time and the y coordinate represents the location at a given time, is called a *one – dimensional random walk*. If we define the random process $Y(t) = Y_n, \ n \leq t < n+1$, then the above figure shows an example of the sample path of $Y(t)$, where the length of each step is s. It is a staircase with discontinuities at $t = kT, k = 1, 2, \ldots$

Suppose that at the end of the n^{th} trial there are exactly k success. Then there are k steps to the right and $n-k$ steps to the left. Thus,

$$Y(nT) = ks - (n-k)s = (2k-n)s = rs$$

where $r = 2k - n$. This implies that $Y(nT)$ is a random variable that assumes values rs, where $r = n, n-2, n-4, \ldots, -n$. Since the event $\{Y(nT) = rs\}$ is the event $\{k \ successes \ in \ n \ trials\}$, where $k = \frac{(n+r)}{2}$, we have that

$$P[Y(nT) = rs] = P\left[\frac{n+r}{2} successes \ in \ n \ trials\right] = \binom{n}{\frac{n+r}{2}} p^{\frac{n+r}{2}} (1-p)^{\frac{n-r}{2}}$$

Note that $(n+r)$ must be an even number. Also, since $Y(nT)$ is the sum of n independent Bernoulli random variables, its mean and variance are given as follows:

$$E[Y(nT)] = nE[X_k] = n[ps - (1-p)s] = (2p-1)ns$$

$$E[X_k^2] = ps^2 + (1-p)s^2 = s^2$$

$$Var[Y(nT)] = nVar[X_k] = n[s^2 - s^2(2p-1)^2] = 4p(1-p)ns^2$$

In the special case where $p = \frac{1}{2}, \ E[Y(nT)] = 0$, and $Var[Y(nT)] = ns^2$.

Gambler's Ruin

The random walk described above assumes that the process can continue forever; in other words, it is unbounded. If the walk is bounded, then the ends of the walk are called **barriers**. These barriers can impose different characteristics on the process. For example, they can be ***reflecting barriers***, which means that on hitting them the walk turns around and continuous. They can also be ***absorbing barriers***, which means that the walk ends.

Consider the following random walk with absorbing barriers, which is generally referred to as the **gambler's ruin**. Suppose a gambler plays a sequence of independent games against an opponent. He starts out with $Rs\ k$, and in each game he wins $Rs\ 1$ with probability p and loses $Rs\ 1$ with probability $q = 1 - p$. When $p > q$, the game is advantageous to the gambler either because he is more skilled than his opponent or the rules of the game favor him. If $p = q$, the game is fair; and if $p < q$, the game is disadvantageous to the gambler.

Assume that he gambler stops when he has a total of $Rs\ N$, which means he has additional $Rs\ (N - k)$ over his initial $Rs\ k$. (Another way to express this is that he plays against an opponent who starts out with $Rs\ (N - k)$ and the game stops when either player has lost all of his or her money.) We are interested in computing the probability r_k that the player will be ruined (or he has lost all of his or her money) after starting with $Rs\ k$.

To solve the problem, we note that at the end of the first game, the player will have the sum of $Rs\ (k + 1)$ if he wins the game (with probability p) and the sum of $Rs\ (k - 1)$ if he loses the game (with probability q). Thus, if he wins the first game, the probability that he will eventually be ruined is r_{k+1}; and if he loses his first game, the probability that he will be ruined is r_{k-1}. There are two boundary conditions in this problem. First $r_0 = 1$, since he cannot gamble when he has no money. Second $r_N = 0$, since he cannot be ruined. Thus, we obtain the following:

$$r_k = qr_{k-1} + pr_{k+1} \quad 0 < k < N$$

Since $p + q = 1$, we obtain

$$(p + q)r_k = qr_{k-1} + pr_{k+1} \quad 0 < k < N$$

and we can write it as

$$p(r_{k+1} - r_k) = q(r_k - r_{k-1})$$

From this we obtain the following:

$$r_{k+1} - r_k = \frac{q}{p}(r_k - r_{k-1}) \quad 0 < k < N$$

Notice that $r_2 - r_1 = \frac{q}{p}(r_1 - r_0) = \frac{q}{p}(r_1 - 1)$,

$r_3 - r_2 = \frac{q}{p}(r_2 - r_1) = \left(\frac{q}{p}\right)^2 (r_1 - 1)$, and so on, we obtain the following:

$$r_{k+1} - r_k = \left(\frac{q}{p}\right)^k (r_1 - 1) \quad 0 < k < N$$

Now,

$$r_k - 1 = r_k - r_0 = (r_k - r_{k-1}) + (r_{k-1} - r_{k-2}) + \cdots + (r_1 - 1)$$

$$= \left[\left(\frac{q}{p}\right)^{k-1} + \left(\frac{q}{p}\right)^{k-2} + \cdots + 1\right](r_1 - 1)$$

Thus, $r_k - 1 = \begin{cases} \frac{1-\left(\frac{q}{p}\right)^k}{1-\left(\frac{q}{p}\right)} (r_1 - 1)\,, & p \neq q \\ k(r_1 - 1) & , \quad p = q \end{cases}$

Recalling the boundary condition that $r_N = 0$ implies that

$$r_1 = \begin{cases} 1 - \frac{1-\left(\frac{q}{p}\right)}{1-\left(\frac{q}{p}\right)^N} & , \quad p \neq q \\ 1 - \frac{1}{N} & , \quad p = q \end{cases}$$

Thus, $r_k = \begin{cases} 1 - \frac{1-\left(\frac{q}{p}\right)^k}{1-\left(\frac{q}{p}\right)^N} = \frac{\left(\frac{q}{p}\right)^k - \left(\frac{q}{p}\right)^N}{1-\left(\frac{q}{p}\right)^N} & , \ p \neq q \\ 1 - \frac{k}{N} & , \quad p = q \end{cases}$

Example 1: A certain student wanted to travel during a break to visit his parents. The bus fare was $Rs\ 20$, but the student had only $Rs\ 10$. He figured out that there was a bar near by where people play card games for money. The student signed up for one where he could bet $Rs\ 1$ per game. If he won the game, he would gain $Rs\ 1$; but if he lost the game, he would lose his $Rs\ 1$ bet. If the probability that he won a game is 0.6 independent of other games, what is the probability that he was not able to make the trip?

Solution: We have, $k = 10$ and $N = 20$. Define $a = \frac{q}{p}$,where $p = 0.6$ and $q = 1 - p = 0.4$. Thus, $a = \frac{2}{3}$ and the probability that he was not able to make the trip is the probability that he was ruined given that he started with $k = 10$.

This is r_{10}, which is given by

$$r_{10} = \frac{\left(\frac{q}{p}\right)^{10} - \left(\frac{q}{p}\right)^{20}}{1 - \left(\frac{q}{p}\right)^{20}} = \frac{\left(\frac{2}{3}\right)^{10} - \left(\frac{2}{3}\right)^{20}}{1 - \left(\frac{2}{3}\right)^{20}} = 0.0170$$

Thus, there is only a very small probability that he will not make the trip.

Semi Random and Random Telegraph signal process

If $N(t)$ represents the number of occurrences of a specific event in $(0, t)$ and $X(t) = (-1)^{N(t)}$, then $\{X(t)\}$ is called a **semi – random telegraph signal process**.

If $\{X(t)\}$ is a semi – random telegraph signal process, α is a r.v which is independent of $X(t)$ and which assumes the values $+1$ and -1 with equal probability and $Y(t) = \alpha X(t)$, then $\{Y(t)\}$ is called a **random telegraph signal process.**

A semi-random telegraph signal process is evolutionary.

It will be proved Module 6.1 that the distribution of $N(t)$ is Poison with mean λt, where the probability of exactly one occurrence in a small interval of length h is λh.

In other words, the process $\{N(t)\}$ is a **Poisson process** with the probability law.

$$P\{N(t)=r\}=\frac{e^{-\lambda t}(\lambda t)^r}{r!};\quad r=0,1,2,\dots$$

If $\{X(t)\}$ is the semi – random telegraph signal process, then as per the definition given above, $X(t)$ can take the values $+1$ and -1 only.

$$P\{X(t)=1\}=P\{N(t)\ is\ even\}$$

$$=P\{N(t)=0\}+P\{N(t)=2\}+P\{N(t)=4\}+\cdots+\cdots$$

(since the events are mutually exclusive)

$$=e^{-\lambda t}\left\{1+\frac{(\lambda t)^2}{2!}+\frac{(\lambda t)^4}{4!}+\cdots+\cdots\right\}$$

$$=e^{-\lambda t}\cos h\,\lambda t$$

$$P\{X(t)=-1\}=P\{N(t)\ is\ odd\}$$

$$=P\{N(t)=1\}+P\{N(t)=3\}+\cdots+\infty$$

(since the events are mutually exclusive)

$$=e^{-\lambda t}\left\{\frac{\lambda t}{1!}+\frac{(\lambda t)^3}{3!}+\cdots+\cdots\right\}$$

$$=e^{-\lambda t}\sin h\,\lambda t$$

$$\therefore\qquad E\{X(t)\}=e^{-\lambda t}(\cos h\,\lambda t-\sin h\lambda\,t)$$

$$=e^{-\lambda t}e^{-\lambda t}=e^{-2\lambda t}$$

Note that $E\{X(t)\}$ is not constant and it is a function of t.

To find $E\{X(t_1)X(t_2)\}$, we required the joint probability distribution of $\{X(t_1),X(t_2)\}$.

Now $P\{X(t_1)=1,X(t_2)=1\}=P\{X(t_1)=1|X(t_2)=1\}P\{X(t_2)=1\}$

$=P\{$even number of occurrences of the event in $(t_1-t_2)\}P\{X(t_2)=1\}$

$=e^{-\lambda\tau}\cos h\,\lambda\tau\times e^{-\lambda t_2}\cos h\,\lambda t_2$; where $\tau=t_1-t_2$

Similarly, $P\{X(t_1) = -1,\ X(t_2) = -1\}$

$$= e^{-\lambda\tau} \cos h\, \lambda\tau e^{-\lambda t_2} \sin h\, \lambda t_2$$

$P\{X(t_1) = 1,\ X(t_2) = -1\} = e^{-\lambda\tau} \sin h\lambda\tau\, e^{-\lambda t_2} \sin h\, \lambda t_2$

and $P\{X(t_1) = -1,\ X(t_2) = 1\} = e^{-\lambda\tau} \sin h\lambda\tau\, e^{-\lambda t_2} \sin h\, \lambda t_2$

Now $X(t_1)X(t_2) = 1$, if $\{X(t_1) = 1\ and\ X(t_2) = 1\}$ or

$$\{X(t_1) = -1\ \ and\ \ X(t_2) = -1\}$$

$\therefore\quad P\{X(t_1)X(t_2) = 1\} = e^{-\lambda(\tau+t_2)} \cos h\lambda\tau\, (\cos h\, \lambda t_2 + \sin h\, \lambda t_2)$

$$= e^{-\lambda\tau} \cos h\, \lambda\tau$$

and $P\{X(t_1)X(t_2) = -1\} = e^{-\lambda(\tau+t_2)} \sin h\, \lambda\tau(\cos h\, t_2 + \sin h\, \lambda t_2)$

$$= e^{-\lambda\tau} \sin h\, \lambda\tau$$

$\therefore R(t_1, t_2) = E\{X(t_1)X(t_2)\} = 1 \times e^{-\lambda\tau} \cos h\, \lambda\tau - 1 \times e^{-\lambda\tau} \sin h\, \lambda\tau = e^{-2\lambda\tau}$

$$= e^{-2\lambda(t_1 - t_2)}$$

Although $R(t_1, t_2)$ is a function of $(t_1 - t_2)$, $E\{X(t)\}$ is not a constant.

Therefore, $\{X(t)\}$ **is evolutionary**.

A random telegraph signal processes is a WSS process.

Let us now consider the random telegraph signal process $\{Y(t)\}$, where
$Y(t) = \alpha X(t)$.
By definition, $P(\alpha = 1) = \frac{1}{2}$ and $P(\alpha = -1) = \frac{1}{2}$

$\therefore\qquad E(\alpha) = 0$ and $E(\alpha^2) = 1$

Now $E\{Y(t)\} = E(\alpha) \times E\{X(t)\} = 0$ [since α and $X(t)$ are independent]

$E\{Y(t_1) \times Y(t_2)\} = E\{\alpha^2 X(t_1) \times X(t_2)\}$

$$= E(\alpha^2)E\{X(t_1)X(t_2)\} \qquad \text{(by independence)}$$

$$= e^{-2\lambda(t_1 - t_2)}$$

$i.e.$, $R_{yy}(t_1, t_2) =$ a function of $(t_1 - t_2)$. Therefore, $\{Y(t)\}$ is a wide – sense stationary process.

Unit-6

Special Stochastic Processes

6.1

Poisson Process

There are many practical situations where the random times of occurrences of some specific events are of primary interest. For example, we may want to study the times at which components fail in a large system or the times at which jobs enter the queue in a computer system or the times of arrival of phone calls at an exchange or the times of emission of electrons from the cathode of a vacuum tube. In these examples, our main interest may not be the event itself but the sequence of random time instants at which the events occur. An ensemble of discrete sets of points from the time domain called a **point process** is used to model and analyse phenomena such as the ones mentioned above. An independent increments point process, $i.e.$, a point process with the property that the number of occurrences in any finite collection of non overlapping time intervals are independent r.vs, leads to a Poisson process.

Definition: If $X(t)$ represents the number of occurrences of a certain event in $(0, t)$, then the discrete random process $\{X(t)\}$ is called the ***Poisson process***, provided the following postulates are satisfied:

(i) $P[1 \text{ occurrence in } (t, t+\Delta t)] = \lambda\Delta t + o(\Delta t)$

(ii) $P[0 \text{ occurrence in } (t, t+\Delta t)] = 1 - \lambda\Delta t + o(\Delta t)$

(iii) $P[2 \text{ or more occurrences in } (t, 1+\Delta t) = o(\Delta t)$

(iv) $X(t)$ is independent of the number of occurrences of the event in any interval prior and after the interval $(0, t)$.

(v) The probability that the event occurs a specified number of times in $(t_0, t_0 + t)$ depends only on t, but not on t_0.

Probability Law for the Poisson Process $\{X(t)\}$

Let λ be the number of occurrences of the event in unit time.

Let $P_n(t) = P\{X(t) = n\}$

$\therefore\ P_n(t + \Delta t) = P\{X(t + \Delta t) = n\}$

$= P\{(n-1)$ calls in $(0, t)$ and 1 call in $(t, t + \Delta t)\}$

$+P\ \{n$ calls in $(0, t)$ and no call in $(t, t + \Delta t)\}$

$= P_{n-1}(t)\lambda\Delta t + P_n(t)(1 - \Delta t)$ (by the postulates (i) and (ii))

$$\therefore \quad \frac{P_n(t+\Delta t) - P_n(t)}{\Delta t} = \lambda\{P_{n-1}(t) - P_n(t)\}$$

Taking the limits as $\Delta t \longrightarrow 0$

$$\frac{d}{dt}P_n(t) = \lambda\{P_{n-1}(t) - P_n(t)\} \quad \text{... (1)}$$

Let the solution of the equation (1) be

$$P_n(t) = \frac{(\lambda t)^n}{n!}f(t) \quad \text{... (2)}$$

Differentiating (2) with respect to t,

$$P_n'(t) = \frac{\lambda^n}{n!}\{nt^{n-1}f(t) + t^n f'(t)\} \quad \text{... (3)}$$

Using (2) and (3) in (1),

$$\frac{\lambda^n}{n!}t^n f'(t) = -\lambda\frac{(\lambda t)^n}{n!}f(t)$$

$i.e.,\qquad f'(t) = -\lambda f(t)$

Integrating, $\quad f(t) = ke^{-\lambda t}$ ... (4)

Taking $n = 0$ in (2) , we get $P_0(t) = f(t)\ \ \forall\ t$

$\therefore\ f(0) = P_0(0)$

$= P\{X(0) = 0\}$

$= P$ {no event occurs in $(0, 0)$} $= 1$ ... (5)

Using (5) in (4), we get $k = 1$ and hence

$$f(t) = e^{-\lambda t} \qquad \ldots (6)$$

Using (6) in (2),

$$P_n(t) = P\{X(t) = n\} = \frac{e^{-\lambda t}(\lambda t)^n}{n!}, \ \ n = 0, ,1 ,2, \ldots, \ldots$$

Thus the probability distribution of $X(t)$ is the Poisson distribution with parameter λt.

Note: *We have assumed that the rate of occurrence of the event λ is a constant, but it can be function of t also. When λ is a constant, the process is called a* ***homogeneous Poisson process****. Unless specified otherwise, the Poisson process will be assumed homogeneous.*

Second Order Probability Function of a Homogeneous Poisson Process:

$P[X(t_1) = n_1, \ X(t_2) = n_2]$

$= P[X(t_1) = n_1]P[X(t_2) = n_2 | X(t_1) = n_1], t_2 > t_1$

$= P[X(t_1) = n_1]P$ [the event occurs $(n_2 - n_1)$ times in the interval of $(t_2 - t_1)$]

$$= \frac{e^{-\lambda t_1}(\lambda t_1)^{n_1}}{n_1!} \frac{e^{-\lambda(t_2-t_1)}\{\lambda(t_2-t_1)\}^{n_2-n_1}}{(n_2-n_1)!}, \text{ if } n_2 \geq n_1$$

$$= \begin{cases} \dfrac{e^{-\lambda t_2} \lambda^{n_2} \ t_1^{n_1} \ (t_2-t_1)^{n_2-n_1}}{n_1! \ (n_2-n_1)!}, & n_2 \geq n_1 \\ 0, & otherwise \end{cases}$$

Proceeding similarly, we can get the third order probability function as

$P[X(t_1) = n_1, X(t_2) = n_2, \ X(t_3) = n_3]$

$$= \begin{cases} \frac{e^{-\lambda t_3} \lambda^{n_3} t_1^{n_1} (t_2 - t_1)^{n_2 - n_1} (t_3 - t_2)^{n_3 - n_2}}{n_1! (n_2 - n_1)! (n_3 - n_2)!}, & n_3 \geq n_2 \geq n_1 \\ 0, & otherwise \end{cases}$$

Mean and Autocorrelation of the Poisson Process

The probability law of the Poisson process $\{X(t)\}$ is the same as that of a Poisson distribution with parameter λt.

$$E\{X(t)\} = Var\{X(t)\} = \lambda t$$

$$\therefore \ \lambda t = Var\{X(t)\} = E(\{X^2(t)\}) - E(\{X(t)\})^2$$

$$\lambda t = E(\{X^2(t)\}) - \lambda^2 t^2$$

$$\therefore \ E\{X^2(t)\} = \lambda t + \lambda^2 t^2 \qquad \dots (1)$$

$$R_{xx}(t_1, t_2) = E\{X(t_1)X(t_2)\}$$

$$= E[X(t_1)\{X(t_2) - X(t_1) + X(t_1)\}]$$

$$= E[X(t_1)\{X(t_2) - X(t_1)\}] + E\{X^2(t_1)\}$$

$$= E[X(t_1)] \ E[X(t_2) - X(t_1)] + E\{X^2(t_1)\}$$

since $\{X(t)\}$ is a process of independent increments.

$$= \lambda t_1 \lambda(t_2 - t_1) + \lambda t_1 + \lambda^2 t_1^2, \text{ if } t_2 \geq t_1 \qquad \text{[by (1)]}$$

$$= \lambda^2 t_1 t_2 + \lambda t_1, \text{ if } t_2 \geq t_1$$

or $R_{xx}(t_1, t_2) = \lambda^2 t_1 t_2 + \lambda \, min(t_1, t_2)$

$$C_{xx}(t_1, t_2) = R_{xx}(t_1, t_2) - E\{X(t_1\} E\{X(t_2)\}$$

$$= \lambda^2 t_1 t_2 + \lambda t_1 - \lambda^2 t_1 t_2$$

$$= \lambda t_1, \text{ if } t_2 \geq t_1$$

or $\qquad = \lambda \min(t_1, t_2)$

$$r_{xx}(t_1, t_2) = \frac{C_{xx}(t_1, t_2)}{\sqrt{Var\{X(t_1)\}}\sqrt{Var\{X(t_2)\}}} = \frac{\lambda t_1}{\sqrt{\lambda t_1 \lambda t_2}} = \sqrt{\frac{t_1}{t_2}}, \text{ if } t_2 \geq t_1$$

Note: *Poisson process is not a stationary process.*

Properties of Poisson Process:

1. The Poisson process is a **Markov process**.

Proof: Consider $P[X(t_3) = n_3 | X(t_2) = n_2,\ X(t_1) = n_1]$

$$= \frac{P[X(t_1)=n_1, X(t_2)=n_2, X(t_3)=n_3]}{P[X(t_1)=n_1, X(t_2)=n_2]}$$

$$= \frac{e^{-\lambda(t_3-t_2)} \lambda^{n_3-n_2} (t_3-t_2)^{n_3-n_2}}{(n_3-n_2)!}$$

[Refer to the second and third order probability functions of the Poisson process]

$$= P[X(t_3) = n_3 | X(t_2) = n_2]$$

This means that the conditional probability distribution of $X(t_3)$ given all the past values $X(t_1) = n_1, X(t_2) = n_2$ depends only on the most recent value $X(t_2) = n_2$.

That is, the Poisson process possesses the Markov property. Hence the result.

2. **Additive Property :**
   Sum of two independent Poisson processes is a Poisson process.

   (**See P1** for proof)

3. Difference of two independent Poisson processes is not a Poisson process

   (**See P2** for proof)

4. The interarrival time of a Poisson process $i.e.$, interval between two successive occurrences of a Poisson process with parameter λ has an exponential distribution with mean $\frac{1}{\lambda}$.

   (**See P3** for proof)

5. If the number of occurrences of an event E in an interval of length t is a Poisson process $\{X(t)\}$ with parameter λt and if each occurrence of E has a constant probability p of being recorded and the recordings are independent of each other then the number $N(t)$ of the recorded occurrences in t is also a Poisson process with parameter λpt.

   (**See P4** for proof)

Example 1: Suppose that customers arrive at a bank according to a Poisson process with a mean rate of 3 per minute; find the probability during a time interval of 2 min (i) exactly 4 customers arrive and (ii) more than 4 customers arrive.

Solution: Mean of the Poisson process $= \lambda t$

Mean arrival rate $=$ mean number of arrivals per minute (unit time) $= \lambda$

Given $\lambda = 3$. We have

$$P\{X(t) = k\} = \frac{e^{-\lambda t}(\lambda t)^k}{k!}$$

$$\therefore \qquad P\{X(2) = 4\} = \frac{e^{-6}6^4}{4!} = 0.133$$

$$P\{X(2) > 4\} = 1 - [P\{X(2) = 0\} + P\{X(2) = 1\} + P\{X(2) = 2\} + P\{X(2) = 3\} + P\{X(2) = 4\}]$$

$$= 1 - \sum_{k=0}^{4} \frac{e^{-6}6^k}{k!} = 0.715$$

Example 2: A machine goes out of order, whenever a component fails. The failure of this part follows a Poisson process with a mean rate of 1 per week. Find the probability that 2 weeks have elapsed since last failure. If there are 5 spare parts of this component in an inventory and that the next supply is not due in 10 weeks, find the probability that the machine will not be out of order in the next 10 weeks.

Solution:

(i) Here the unit times is 1 week

Mean failure rate $=$ mean number of failure in a week $= \lambda = 1$.

P {no failures in the 2 weeks since last failure} $= P\{X(2) = 0\}$

$$= \frac{e^{-2\lambda}(2\lambda)^0}{0!} = e^{-2} = 0.135$$

(ii) There are only 5 spare parts and the machine should not go out of order in the next 10 weeks.

P{for this event} $= P\{X(10) \leq 5\}$

$$= \sum_{k=0}^{5} \frac{e^{-10}10^k}{k!} = 0.068$$

Example 3: If $\{N_1(t)\}$ and $\{N_2(t)\}$ are 2 independent Poisson processes with parameters λ_1 and λ_2 respectively, show that

$$\boldsymbol{P[N_1(t) = k|\{N_1(t) + N_2(t) = n\}] = {}^nC_k\, p^k q^{n-k},}$$ **where**

$$\boldsymbol{p = \frac{\lambda_1}{\lambda_1+\lambda_2}}$$ **and** $$\boldsymbol{q = \frac{\lambda_2}{\lambda_1+\lambda_2}}$$

Solution: Required conditional probability

$$= \frac{P[\{N_1(t)=k\} \cap \{N_1(t)+N_2(t)=n\}]}{P\{N_1(t)+N_2(t)=n\}}$$

$$= \frac{P[\{N_1(t)=k\} \cap \{N_2(t)=n-k\}]}{P\{N_1(t)+N_2(t)=n\}}$$

$$= \frac{\frac{e^{-\lambda_1 t}(\lambda_1 t)^k}{k!} \frac{e^{-\lambda_2 t}(\lambda_2 t)^{n-k}}{(n-k)!}}{\frac{e^{-(\lambda_1+\lambda_2)t}\{(\lambda_1+\lambda_2)t\}^n}{n!}}$$

(by independence and additive property)

$$= \frac{n!}{k!(n-k)!} \frac{(\lambda_1 t)^k(\lambda_2 t)^{n-k}}{\{(\lambda_1+\lambda_2)t\}^n}$$

$$= {}^nC_k \left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k \left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-k}$$

$$= {}^nC_k\, p^k q^{n-k}$$

Example 4: **If customers arrive at a counter in a accordance with a Poison process with a mean rate of 2 per minute, find the probability that the interval between 2 consecutive arrivals is (i) more than 1 min, (ii) between 1 min and 2 min and (iii) 4 min or less.**

Solution: Refer to property 4 of Poisson process.

The interval T between 2 consecutive arrivals follows an exponential distribution with parameter $\lambda = 2$.

(i) $P(T > 1) = \int_1^\infty 2e^{-2t} dt = e^{-2} = 0.135$

(ii) $P(1 < T < 2) = \int_1^2 2e^{-2t} dt = e^{-1} - e^{-2} = 233$

(iii) $P(T \le 4) = \int_0^4 2e^{-2t} dt = 1 - e^{-8} = 0.999$

Example 5: A radioactive source emits particles at a rate of 5 per minute in accordance with Poisson process. Each particle emitted has a probability $\mathbf{0.6}$ of being recorded. Find the probability that $\mathbf{10}$ particles are recorded in $\boldsymbol{4 - min}$ period.

Solution: Refer to property 5 Poisson processes.
The number of recorded particles $N(t)$ follows a Poisson process with parameter λp. Here $\lambda = 5$ and $p = 0.6$

$$\therefore P\{N(t) = k\} = \frac{e^{-\lambda pt}(\lambda pt)^k}{k!} = \frac{e^{-3t}(3t)^k}{k!}$$

$$\therefore P\{N(4) = 10\} = \frac{e^{-12}(12)^{10}}{10!} = 0.104$$

Example 6: The number of accidents in a city follows a Poisson process with a mean of 2 per day and the number $\boldsymbol{X_i}$ of people involved in the $\boldsymbol{i^{th}}$ accident has the distribution (independent) $\boldsymbol{P\{X_i = k\} = \frac{1}{2^k}(k \geq 1)}$. Find the mean and variance of the number of people involved in accidents per week.

Solution: The mean and variance of the distribution

$P\{X_i = k\} = \frac{1}{2^k}$, $k = 1, 2, 3, \dots, \dots$ can be obtained as 2 and 2.

Let the number of accidents on any day be assumed as n.

The number of people involved in these accidents be $X_1, X_2, \dots X_n$.

$X_1, X_2, \dots X_n$ are independent and identically distributed $r.vs$ with mean 2 and variance 2.

Therefore, by central limit theorem, $(X_1 + X_2 + \cdots + X_n)$ follows a normal distribution with mean $2n$ and variance $2n$, $i.e.$, the total number of people involved in all the accidents on a day with n accidents is $2n$.

If N denotes number of people involved in accidents on any day, then

$P\{N = 2n\} = P\{X(t) = n\}$ [where $X(t)$ is the number of accidents]

$$= \frac{e^{-2t}(2t)^n}{n!} \text{ (by data)}$$

$$\therefore \qquad E(N) = \sum_{n=0}^{\infty} \frac{2ne^{-2t}(2t)^n}{n!}$$

$$= 2E\{X(t)\} = 4t$$

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

$$= \sum_{n=0}^{\infty} \frac{4n^2e^{-2t}(2t)^n}{n!} - 16t^2$$

$$= 4E\{X^2(t)\} - 16t^2$$

$$= 4[Var\{X(t)\} + E^2\{X(t)\}] - 16t^2$$

$$= 4[2t + 4t^2] - 16t^2 = 8t$$

Therefore, mean and variance of the number of people involved in accidents per week are 28 and 56 respectively.

Example 7: If T_n is the r.v denoting the time of occurrence of the n^{th} event in a Poisson process with parameter λ, show that the distribution function $F_n(t)$ of T_n is given by

$$F_n(t) = \begin{cases} 1 - \sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t} & , \quad if \quad t \geq 0 \\ 0 & , \quad if \quad t < 0 \end{cases}$$

Deduce the density function $f_n(t)$ of T_n

Solution:

$$F_n(t) = P\{T_n \leq t\} = 1 - P\{T_n > t\}$$

when $T_n > t$, $i.e.$, the time of occurrence of the n^{th} event $> t$, $(n-1)$ or less events must have occurred in $(0, t)$

$$\therefore \quad F_n(t) = 1 - P\{X(t) \leq n-1\}$$

$$= 1 - \sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \text{ when } t \geq 0$$

Differentiating both sides with respect to t and noting that $F_n'(t) = f_n(t)$

$$f_n(t) = -\sum_{k=0}^{n-1} \left\{ -\lambda \frac{(\lambda t)^k}{k!} e^{-\lambda t} + \lambda \frac{(\lambda t)^{k-1}}{(k-1)!} e^{-\lambda t} \right\}$$

$$= \lambda e^{-\lambda t} \sum_{k=0}^{n-1} \left\{ \frac{(\lambda t)^k}{k!} e^{-\lambda t} - \frac{(\lambda t)^{k-1}}{(k-1)!} \right\}$$

$$= \lambda e^{-\lambda t} \left[1 + \left\{ \frac{\lambda t}{1!} - 1 \right\} + \left\{ \frac{(\lambda t)^2}{2!} - \frac{\lambda t}{1!} \right\} + \ldots + \left\{ \frac{(\lambda t)^{n-1}}{(n-1)!} - \frac{(\lambda t)^{n-2}}{(n-2)!} \right\} \right]$$

$$= \frac{\lambda^n t^{n-1} e^{\lambda t}}{(n-1)!}, \quad t \geq 0$$

Example 8 : If $\{X(t)\}$ is a Poisson process, prove that

$$\boldsymbol{P\{X(s) = r | X(t) = n\} = \binom{n}{r} \left(\frac{s}{t}\right)^r \left(1 - \frac{s}{t}\right)^{n-1}} \textbf{ where } \boldsymbol{s} < t$$

Solution:

$$P\{X(s) = r | X(t) = n\} = \frac{P[\{X(s)=r\} \cap \{X(t)=n\}]}{P\{X(t)=n\}}$$

$$= \frac{P\{X(s)=r \cap X(t-s)=n-r\}}{P\{X(t)=n\}}$$

$$= \frac{P\{X(s)=r\}P\{X(t-s)=n-r\}}{P\{X(t)=n\}} \quad \text{(by independence)}$$

$$= \frac{\dfrac{e^{-\lambda s}(\lambda s)^r}{r!}\,\dfrac{e^{-\lambda(t-s)}\left[\lambda(t-s)^{n-r}\right]}{(n-r)!}}{\dfrac{e^{-\lambda t}(\lambda t)^n}{n!}}$$

$$= \frac{n!}{\boldsymbol{r!}\,(n-r)!}\frac{s^r(t-s)^{n-r}}{t^n} = {}^nC_r\left(\frac{s}{t}\right)^r\left(1-\frac{s}{t}\right)^{n-r}$$

6.2

Gaussian Process

Many random phenomena in physical problems including *noise* are well approximated by a special class of random process, namely Gaussian random process. A number of processes such as the Wiener process and the shot noise process can be approximated, as per central limit theorem, by a Gaussian process. Moreover the output of a linear system in which the input is a weighted sum of a large number of independent samples of a random process tends to approach a Gaussian process. Gaussian processes play an important role in the theory and analysis of random phenomena, because they are good approximations to the observations and multivariate Gaussian distributions are analytically simple.

One of the most important uses of the Gaussian process is to model and analyse the effects of thermal noise in electronic circuits used in communication systems. Individual circuits contain resistors, inductors and capacitors as well as semiconductor devices. The resistors and semiconductor elements contain charged particles (free electrons) subjected to random motion due to thermal energy. The random motion of charged particles causes fluctuations in the current waveforms or information bearing signals that flow through these components. These fluctuations are called **thermal noise**, which are of sufficient strength to disturb a weak signal and to make the recognition of signals a difficult task. Models of thermal noise are used to identify and minimize the effects of noise in signal recognition.

Gaussian Process: A real valued random process $\{X(t)\}$ is called a **Gaussian process** or **normal process**, if the random variables $X(t_1), X(t_2), \dots, X(t_n)$ are jointly normal for every $n = 1,2, \dots$ and for any set of t_i's.

The n^{th} order density of a Gaussian process is given by

$$f(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Lambda|^{\frac{1}{2}}} exp\left[-\frac{1}{2|\Lambda|}\sum_{i=1}^{n}\sum_{j=1}^{n}|\Lambda|_{ij}(x_i - \mu_i)(x_j - \mu_j)\right]$$

where $\mu_i = E\{X(t_i)\}$ and Λ is the n^{th} order square matrix (λ_{ij}), where $\lambda_{ij} = C\{X(t_i), X(t_j)\}$ and $|\Lambda|_{ij} =$ cofactor of λ_{ij} in $|\Lambda|$ ... (1)

Note: *Gaussian process is completely specified by the first and second order moments, viz., means and covariances (variances).*

Note: When we consider the first order density of a Gaussian process,

$$\Lambda = (\lambda_{11}) = [cov(X(t_1), X(t_1))] = [Var\{X(t_1)\}] = \sigma_1^2$$

$\therefore |\Lambda| = \sigma_1^2$ and $|\Lambda|_{11} = 1$

$$\therefore f(x_1, t_1) = \frac{1}{\sigma_1\sqrt{2\pi}} exp\left\{-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}\right\}$$

Note: When we consider the second order density of a Gaussian process,

$$\Lambda = \begin{pmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & r_{12}\sigma_1\sigma_2 \\ r_{21}\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

$\therefore |\Lambda| = \sigma_1^2\sigma_2^2(1-r^2)$, where $r_{12} = r_{21} = r$

$|\Lambda|_{11} = \sigma_2^2, |\Lambda|_{12} = -r\sigma_1\sigma_2, |\Lambda|_{21} = -r\sigma_1\sigma_2,\ |\Lambda|_{22} = \sigma_1^2$

$\therefore f(x_1, x_2; t_1, t_2)$

$$= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} exp\left[-\frac{1}{2\sigma_1^2\sigma_2^2(1-r^2)}\{\sigma_2^2(x_1-\mu_1)^2 - 2r\sigma_1\sigma_2(x_1-\mu_1)(x_2-\mu_2) + \sigma_1^2(x_2-\mu_2)^2\}\right]$$

$$i.e., f(x_1, x_2; t_1, t_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} exp\left[-\frac{1}{2(1-r^2)}\left\{\frac{(x_1-\mu_1)^2}{\sigma_1^2} - \frac{2r(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right\}\right]$$

Properties

1. If a Gaussian process is wide sense stationary, it is also strict sense stationary. (**See P1** for proof).
2. If the member functions of a Gaussian process are uncorrelated, then they are independent. (**See P2** for proof).

3. If the input $\{X(t)\}$ of a linear system is a Gaussian process, the output will also be a Gaussian process.

Example 1: If $\{X(t)\}$ is a Gaussian process with $\mu(t) = 10$ and $C(t_1, t_2) = 16e^{-|t_1 - t_2|}$, find the probability that

(i) $X(10) \leq 8$ and (ii) $|X(10) - X(6)| \leq 4$.

Solution: Since $\{X(t)\}$ is a Gaussian process, any member of the process is a normal r.v.

Therefore, $X(10)$ is a normal r.v with mean $\mu(10) = 10$ and variance $C(10,10) = 16$.

i. $P\{X(10) \leq 8\} = P\left\{\frac{X(10)-10}{4} \leq -0.5\right\}$

$= P\{Z \leq -0.5\}$ (where Z is the standard normal r.v)

$= 0.5 - P\{0 \leq Z \leq 0.5\}$

$= 0.5 - 0.1915$ (from normal tables)

$= 0.3085$

ii. $X(10) - X(6)$ is also a normal r.v with mean

$\mu(10) - \mu(6) = 10 - 10 = 0.$

$$Var\{X(10) - X(6)\} = Var\{X(10)\} + Var\{X(6)\} - 2Cov\{X(10), X(6)\}$$

$$= C(10,10) + C(6,6) - 2C(10,6)$$

$$= 16 + 16 - 2.16e^{-4}$$

$$= 31.4139$$

$$P\{X(10) - X(6) \leq 4\} = P\left\{\frac{|X(10)-X(6)|}{5.6048} \leq \frac{4}{5.6048}\right\}$$

$= P\{|Z| \le 0.7137\}$

$= 2 \times 0.2611$

$= 0.5222$

Example 2: The process $\{X(t)\}$ is normal with $\mu_t = 0$ and $R_x(\tau) = 4e^{-3|\tau|}$. Find a memoryless system $g(x)$ such that the first order density $f_y(y)$ of the resulting output $Y(t) = g\{X(t)\}$ is uniform in the interval $(6, 9)$.

Solution: Since $\{X(t)\}$ is a normal process, a sample function $X(t)$ follows a normal distribution with mean 0 and variance given by $R_x(0) = 4$.

$$\therefore f_X(x) = \frac{1}{2\sqrt{2\pi}} e^{-\frac{x^2}{8}}, -\infty < x < \infty$$

Now $Y(t)$ is to be uniform in $(6,9)$

$$\therefore f_Y(y) = \frac{1}{3}, 6 < y < 9$$

Therefore, the distribution function of Y is given by

$$F_Y(y) = \int_6^y f_Y(y)dy = \frac{1}{3}(y-6) \qquad \text{.....(1)}$$

Now $F_Y(y) = P\{Y(t) \le y\} = P\{g[X(t)] \le y\}$

$= P\{X(t) \le g^{-1}(y)\}$

$= P\{X(t) \le x\}$ [since $y = g(x)$]

$= F_X(x)$

But, from (1), $F_Y\{g(x)\} = \frac{1}{3}\{g(x) - 6\}$

$$\therefore \frac{1}{3}\{g(x) - 6\} = F_X(x)$$

$$\therefore g(x) = 6 + 3F_X(x) = 6 + 3\int_{-\infty}^{x} \frac{1}{2\sqrt{2\pi}} e^{-\frac{x^2}{8}} dx$$

Example 3: It is given that $R_x(\tau) = e^{-|\tau|}$ for a certain stationary Gaussian random process $\{X(t)\}$. Find the joint p.d.f of the r.vs $X(t), X(t+1), X(t+2)$.

Solution: Let us denote the given r.vs by $X(t_1), X(t_2), X(t_3)$.

The joint p.d.f of $\{X(t_1), X(t_2), X(t_3)\}$ is given by

$$f(x_1, x_2, x_3; t_1, t_2, t_3) = \frac{1}{(2\pi)^{\frac{3}{2}}|\Lambda|^{\frac{1}{2}}} exp\left[-\frac{1}{2|\Lambda|}\sum_{i=1}^{3}\sum_{j=1}^{3}|\Lambda|_{ij}(x_i - \mu_i)\left(x_j - \mu_j\right)\right]$$

where $\mu_i = E\{X(t_i)\}$ and Λ is the third order square matrix (λ_{ij}), where $\lambda_{ij} = C\{X(t_i), X(t_j)\}$ and $|\Lambda|_{ij}$ = cofactor of λ_{ij} in $|\Lambda|$.

$$E\{X(t)\} = \sqrt{\lim_{\tau \to \infty} R_x(\tau)} = \sqrt{\lim_{\tau \to \infty} e^{-|\tau|}} = 0$$

$$\therefore \lambda_{ij} = C\{X(t_i)X(t_j)\} = R(t_i - t_j) \quad (\because \text{it is stationary})$$

$$\therefore \lambda_{11} = R(0) = 1, \lambda_{12} = R(1) = e^{-1} \text{ etc.} \qquad \text{(Compute!)}$$

$$\therefore \Lambda = \begin{pmatrix} 1 & \frac{1}{e} & \frac{1}{e^2} \\ \frac{1}{e} & 1 & \frac{1}{e} \\ \frac{1}{e^2} & \frac{1}{e} & 1 \end{pmatrix} \text{ and } |\Lambda| = \left(1 - \frac{1}{e^2}\right)^2$$

$$|\Lambda|_{11} = 1 - \frac{1}{e^2}, |\Lambda|_{12} = -\frac{1}{e} + \frac{1}{e^3}, |\Lambda|_{13} = 0 \text{ etc.} \qquad \text{(do it !)}$$

Therefore, the required joint p.d.f is given by

$$= \frac{1}{(2\pi)^{\frac{3}{2}}\left(1 - \frac{1}{e^2}\right)} exp\left[-\frac{1}{2\left(1 - \frac{1}{e^2}\right)^2}\left\{\left(1 - \frac{1}{e^2}\right)x_1^2 - \frac{2}{e}\left(1 - \frac{1}{e^2}\right)x_1x_2 + \left(1 - \frac{1}{e^4}\right)x_2^2 - \frac{2}{e}\left(1 - \frac{1}{e^2}\right)x_2x_3 + \left(1 - \frac{1}{e^2}\right)x_3^2\right\}\right]$$

$i.e.,$

$$= \frac{1}{(2\pi)^{\frac{3}{2}}\left(1-\frac{1}{e^2}\right)} exp\left[-\frac{1}{2\left(1-\frac{1}{e^2}\right)}\left\{x_1^2 - \frac{2}{e}x_1x_2 + \left(1+\frac{1}{e^2}\right){x_2}^2 - \ \frac{2}{e}x_2x_3 + x_3^2\right\}\right]$$

6.3

Processes Depending on Stationary Gaussian Process

Square law detector process

If $\{X(t)\}$ is a zero mean stationary Gaussian process and if $Y(t) = X^2(t)$, then $\{Y(t)\}$ is called a **square law detector process**.

$$E\{Y(t)\} = E\{X^2(t)\} = Var\{X(t)\} = R_{xx}(0)$$

$$R_{yy}(t_1, t_2) = E\{Y(t_1)Y(t_2)\} = E\{X^2(t_1)X^2(t_2)\}$$

$$= E\{X^2(t_1)\}E\{X^2(t_2)\} + 2E^2\{X(t_1)X(t_2)\}$$

[Since X and Y are jointly normal, $E(X^2Y^2) = E(X^2)E(Y^2) + 2E^2(XY)$]

$$= R_{xx}^2(0) + 2R_{xx}^2(\tau) \text{ [since } X(t) \text{ is stationary]}$$

Since the RHS is a function of τ, LHS is also a function of $\tau = t_1 - t_2$.

$i.e.$,

$$R_{yy}(\tau) = R_{xx}^2(0) + 2R_{xx}^2(\tau)$$

Therefore, $\{Y(t)\}$ is also a stationary process (at least in the wide-sense).

We note that $E\{Y^2(t)\} = R_{yy}(0) = 3R_{xx}^2(0)$

$Var\{Y(t)\} = E\{Y^2(t)\} - (E\{Y(t)\})^2 = 3R_{xx}^2(0) - R_{xx}^2(0) = 2R_{xx}^2(0)$ and $C_{yy}(\tau) = 2R_{xx}^2(\tau)$

Power spectral density of $\{Y(t)\}$ is given by

$$S_{yy}(\omega) = \int_{-\infty}^{\infty} R_{yy}(\tau)\, e^{-i\omega\tau}\, d\tau = \int_{-\infty}^{\infty} \{R_{xx}^2(0) + 2R_{xx}^2(\tau)\} e^{-i\tau\omega} d\tau$$

$$= 2\pi R_{xx}^2(0)\delta(\omega) + 2F\{R_{xx}(\tau)R_{xx}(\tau)\} \qquad \dots(1)$$

[since $F^{-1}\{2\pi m^2\delta(\omega)\} = m^2$]

Consider $F^{-1}\{S_{xx}(\omega) * S_{xx}(\omega)\}$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} S_{xx}(\omega - \alpha)S_{xx}(\alpha)e^{i\tau\omega}d\alpha d\omega \qquad \ldots.(2)$$

Put $\omega - \alpha = \beta$ and $\alpha = \gamma$ $i.e.$, $\omega = \beta + \gamma$, $\alpha = \gamma$

Then, from calculus,

$$d\omega d\alpha = \begin{vmatrix} \frac{\partial\omega}{\partial\beta} & \frac{\partial\omega}{\partial\gamma} \\ \frac{\partial\alpha}{\partial\beta} & \frac{\partial\alpha}{\partial\gamma} \end{vmatrix} d\beta d\gamma = \begin{vmatrix} 1 & 1 \\ 0 & 1 \end{vmatrix} d\beta d\gamma = d\beta d\gamma \qquad \ldots(3)$$

Using (3) in (2), we get

$$F^{-1}\{S_{xx}(\omega) * S_{xx}(\omega)\} = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} S_{xx}(\beta)S_{xx}(\gamma)e^{i\tau(\beta+\gamma)}d\beta d\gamma$$

$$= 2\pi\left(\frac{1}{2\pi}\int_{-\infty}^{\infty} S_{xx}(\beta)e^{i\tau\beta}d\beta\right)\left(\frac{1}{2\pi}\int_{-\infty}^{\infty} S_{xx}(\gamma)e^{i\tau\gamma}d\gamma\right)$$

$$= 2\pi R_{xx}(\tau)R_{xx}(\tau) \qquad \ldots(4)$$

Using (4) in (1),

$$S_{yy}(\omega) = 2\pi R^2{}_{xx}(0)\delta(\omega) + \frac{1}{\pi}S_{xx}(\omega) * S_{xx}(\omega)$$

Example 1: If $\{\boldsymbol{Y(t)}\}$ is the square law detector process and if $\boldsymbol{Z(t) = Y(t) - E\{Y(t)\}}$, show that the spectral density of $\{\boldsymbol{Z(t)}\}$ is given by $\boldsymbol{S_{zz}(\omega) = \frac{1}{\pi}\int_{-\infty}^{\infty} S_{xx}(\alpha)S_{xx}(\omega-\alpha)d\alpha}$, where $\boldsymbol{S_{xx}(\omega)}$ is the input spectral density.

Solution:

Note: $Z(t)$ is called the **fluctuation of the square law detector.**

$$E\{Z(t)Z(t-\tau)\} = E[\{Y(t) - E[Y(t)]\}\{Y(t-\tau) - E[Y(t-\tau)]\}]$$

$$= E\{Y(t)Y(t-\tau)\} - E\{Y(t)\}E\{Y(t-\tau)\} \text{ (simplification!)}$$

$$i.e., \; R_{zz}(\tau) = R_{yy}(\tau) - E\{Y(t)\}E\{Y(t-\tau)\}$$

$$= R_{xx}^2(0) + 2R_{xx}^2(\tau) - R_{xx}^2(0)$$

(By square law detector process)

$$= 2R_{xx}^2(\tau)$$

$$\Longrightarrow R_{zz}(\tau) = 2R_{xx}^2(\tau)$$

Taking Fourier transforms,

$$S_{zz}(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} R_{zz}(\tau)e^{-i\omega\tau}d\tau = \frac{1}{2\pi}\int_{-\infty}^{\infty} 2R_{xx}^2(\tau)e^{-i\omega\tau}d\tau$$

$$= \frac{1}{\pi}\int_{-\infty}^{\infty} S_{xx}(\alpha)S_{xx}(\omega-\alpha)d\alpha$$

Two Important Results

We now consider two important results which will be used in the discussion of other processes depending *on stationary Gaussian process*, that will follow.

Result 1: If X and Y are two normal r.vs with zero means, variances σ_1^2 and σ_2^2 and correlation coefficient r, then the probability that they are of the same sign $=\frac{1}{2}+\frac{1}{\pi}sin^{-1}(r)$ and the probability that they are of opposite sign $=\frac{1}{2}-\frac{1}{\pi}sin^{-1}(r)$.

Result 2: If X and Y are two normal r.vs with zero means, variances σ_1^2 and σ_2^2 and correlation coefficient r, then $E\{|XY|\}=\frac{2}{\pi}\sigma_1\sigma_2(\cos\alpha+\alpha\sin\alpha)$, where $\sin\alpha=r$.

Full wave linear detector process

If $\{X(t)\}$ is a zero mean stationary Gaussian process and if $Y(t)=|X(t)|$, then $\{Y(t)\}$ is called **a full wave linear detector process**.

$$E\{Y(t)\}=E\{|X(t)|\}=\int_{-\infty}^{\infty}|x|\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{x^2}{2\sigma^2}}dx$$

$$=\frac{1}{\sigma}\sqrt{\frac{2}{\pi}}\int_{0}^{\infty}xe^{-\frac{x^2}{2\sigma^2}}dx$$

$$=\frac{1}{\sigma}\sqrt{\frac{2}{\pi}}\sigma^2\int_{0}^{\infty}e^{-t}dt\ ,putting\ \frac{x^2}{2\sigma^2}=t$$

$$=\sigma\sqrt{\frac{2}{\pi}}=\sqrt{\frac{2R_{xx}(0)}{\pi}}$$

$$R_{yy}(t_1,t_2)=E\{Y(t_1)Y(t_2)\}=E\{|X(t_1)X(t_2)|\}$$

$$=\frac{2}{\pi}\sigma^2(cos\,\alpha+\alpha\,sin\,\alpha)\quad \text{(using Result 2)}$$

where $\sin\alpha=r=\frac{C\{X(t_1),X(t_2)\}}{\sigma^2}$

$$=\frac{E\{X(t_1)\ X(t_2)\}}{\sigma^2}$$

$$= \frac{R_{xx}(t_1 - t_2)}{\sigma^2} \qquad \text{[since } \{X(t)\} \text{ is stationary]}$$

Therefore, $\{Y(t)\}$ is wide-sense stationary, with

$R_{yy}(\tau) = \frac{2}{\pi} R_{xx}(0)(\cos\alpha + \alpha \sin\alpha)$, where $sin\, \alpha = \frac{R_{xx}(\tau)}{R_{xx}(0)}$

Now $E\{Y^2(t)\} = R_{yy}(0) = \frac{2}{\pi} R_{xx}(0)\left\{0 + \frac{\pi}{2} 1\right\}$,

Since $\sin\alpha = \frac{R_{xx}(0)}{R_{xx}(0)} = 1$ and $\alpha = \frac{\pi}{2}$

$\therefore E\{Y^2(t)\} = R_{xx}(0)$ and $Var\{Y(t)\} = \left(1 - \frac{2}{\pi}\right) R_{xx}(0)$

Half-wave linear detector process

If $\{X(t)\}$ is a zero mean stationary Gaussian process and if

$$Z(t) = \begin{cases} X(t) & , \quad for\ X(t) \geq 0 \\ 0 & , \quad for\ X(t) < 0 \end{cases}$$

then $\{Z(t)\}$ is called a **half-wave linear detector process**.

$Z(t)$ can be rewritten as $Z(t) = \frac{1}{2}\{X(t) + |X(t)|\}$

$\therefore E\{Z(t)\} = \frac{1}{2}[E\{X(t)\} + E\{|X(t)|\}]$

$$= \frac{1}{2}\left[0 + \sqrt{\frac{2}{\pi}} R_{xx}(0)\right] \qquad \text{(refer to the full wave linear detector process)}$$

$$= \sqrt{\frac{R_{xx}(0)}{2\pi}}$$

$$E\{Z(t)Z(t-\tau)\} = E[E\{Z(t)Z(t-\tau)|X(t)X(t-\tau)\}] \qquad \ldots (1)$$

Now $Z(t)Z(t-\tau)|X(t)X(t-\tau) = \frac{1}{2}\{X(t)X(t-\tau) + |X(t)X(t-\tau)|\}$ (or) $= 0$

The first value is assumed, when $X(t)X(t-\tau) > 0$, $i.e.$, when $X(t)$ and $X(t-\tau)$ are both positive or both negative.

$\therefore P\{\text{The first value of assumed}\} = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$

Similarly, $P\{\text{the second value is assumed}\} = \frac{1}{2}$

$$\therefore E\{Z(t)Z(t-\tau) | X(t)X(t-\tau)\} = \frac{1}{4}\{X(t)X(t-\tau) + |X(t)X(t-\tau)|\} \qquad \ldots (2)$$

Using (2) in (1), we get,

$$E\{Z(t)Z(t-\tau)\} = \frac{1}{4}\left\{E\{X(t)X(t-\tau)\} + E\{|X(t)X(t-\tau)|\}\right\}$$

$$= \frac{1}{4}\left\{R_{xx}(\tau) + R_{yy}(\tau)\right\}$$

[where $\{Y(t)\}$ is the full-wave linear detector process]

i.e., $R_{zz}(\tau) = \frac{1}{4}\left[R_{xx}(\tau) + \frac{2}{\pi}R_{xx}(0)(\cos\alpha + \alpha\sin\alpha)\right]$, where $\sin\alpha = \frac{R_{xx}(\tau)}{R_{xx}(0)}$

Therefore, the process $\{Z(t)\}$ is wide-sense stationary.

Now $E\{Z^2(t)\} = R_{zz}(0) = \frac{1}{2}R_{xx}(0)$

$$\therefore Var\{Z(t)\} = E\{Z^2(t)\} - (E\{Z(t)\})^2$$

$$= \frac{1}{2}R_{xx}(0) - \frac{1}{2\pi}R_{xx}(0) = \frac{1}{2}\left(1 - \frac{1}{\pi}\right)R_{xx}(0)$$

Hard limiter process

If $\{X(t)\}$ is a zero mean stationary Gaussian process and if

$$Y(t) = \begin{cases} 1 & for\ X(t) \geq 0 \\ -1 & for\ X(t) < 0 \end{cases}$$

then $\{Y(t)\}$ is called a **hard limiter process** or **ideal limiter process**.

$$E\{Y(t)\} = P\{X(t) \geq 0\} - P\{X(t) < 0\}$$

$$= 0$$

Now

$$Y(t)Y(t-\tau) = \begin{cases} 1, & \text{if } X(t)X(t-\tau) \geq 0 \\ -1, & \text{if } X(t)X(t-\tau) < 0 \end{cases}$$

$i.e., P\{Y(t)Y(t-\tau) = 1\} = P\{X(t)X(t-\tau) \geq 0\}$

$$= \frac{1}{2} + \frac{1}{\pi} sin^{-1}(r_{xx}) \qquad \text{(by Result 1)}$$

$$= \frac{1}{2} + \frac{1}{\pi} sin^{-1}\left\{\frac{R_{xx}(\tau)}{R_{xx}(0)}\right\}$$

and $P\{Y(t)Y(t-\tau) = -1\} = P\{X(t)X(t-\tau) < 0\}$

$$= \frac{1}{2} - \frac{1}{\pi} \sin^{-1}\left\{\frac{R_{xx}(\tau)}{R_{xx}(0)}\right\} \quad \text{(by Result 1)}$$

$$\therefore E\{Y(t)Y(t-\tau)\} = \frac{2}{\pi} \sin^{-1}\left\{\frac{R_{xx}(\tau)}{R_{xx}(0)}\right\}$$

$$i.e., R_{yy}(\tau) = \frac{2}{\pi} \sin^{-1}\left\{\frac{R_{xx}(\tau)}{R_{xx}(0)}\right\} \qquad \ldots.(1)$$

Thus (1) is called the **arcsine law**

Therefore $\{Y(t)\}$ is wide-sense stationary.

Also $E\{Y^2(t)\} = 1$ and $Var\{Y(t)\} = 1$

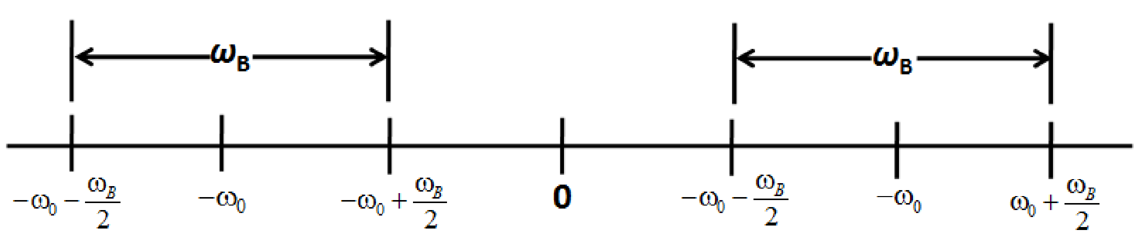
Band Pass Process (Signal)

If the power spectrum of a random process $\{X(t)\}$ is zero outside a certain band (an interval in the ω-axis),

i.e., $S_{xx}(\omega) \neq 0$, in $|\omega - \omega_0| \leq \frac{\omega_B}{2}$ and in $|\omega + \omega_0| \leq \frac{\omega_B}{2}$

and $S_{xx}(\omega) = 0$, in $|\omega - \omega_0| > \frac{\omega_B}{2}$ and in $|\omega + \omega_0| > \frac{\omega_B}{2}$

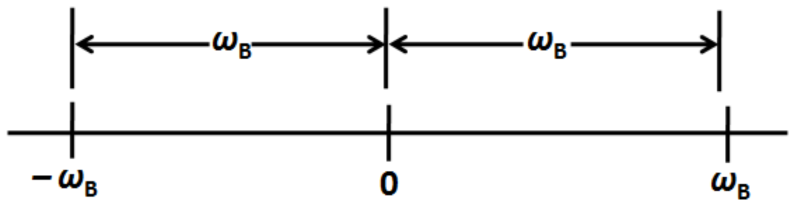
then $\{X(t)\}$ is called a **band pass process**.



If $S_{xx}(\omega) \neq 0$, in $|\omega| \leq \omega_B$ and

$S_{xx}(\omega) \ = 0$, in $|\omega| > \omega_B$

then $\{X(t)\}$ is called a low pass process or ideal low pass process.



If the bandwith ω_B of a band pass process is small compared with the centre frequaency ω_0, the process is called **narrow band process** or **quasimonochromatic.**

If the power spectrum $S_{xx}(\omega)$ of a bandpass process $\{X(t)\}$ is an impulse function, then the process is called **monochromatic**.

Narrow-Band Gaussian Process

In communication system, information bearing signals are often narrow-band Gaussian processes. When such signals are viewed on an oscilloscope, they appear like a sine wave with slowly varying amplitude and phase. Hence a narrow-band Gaussian process $\{X(t)\}$ is often represented as

$$X(t) = R_X(t) \cos[\omega_0 \pm \theta_X(t)] \qquad \dots (1)$$

$R_X(t)$ and $\theta_X(t)$, which are low pass processes, are called the **envelope** and phase of the process $\{X(t)\}$ respectively. (1) can be rewritten as

$$X(t) = [R_X(t) \cos \theta_X(t)] \cos \omega_0 t \mp [R_X(t) \cos \theta_X(t)] \sin \omega_0 t \qquad \dots (2)$$

$R_X(t) \cos \theta_Y(t)$ is called the **inphase component** of the process $\{X(t)\}$ and denoted as $X_c(t)$ or $I(t)$. $R_X(t) \sin \theta_X(t)$ is called the **quadrature component** of $\{X(t)\}$ and denoted as $X_s(t)$ or $Q(t)$.

Both $X_c(t)$ and $X_s(t)$ are low pass processes.

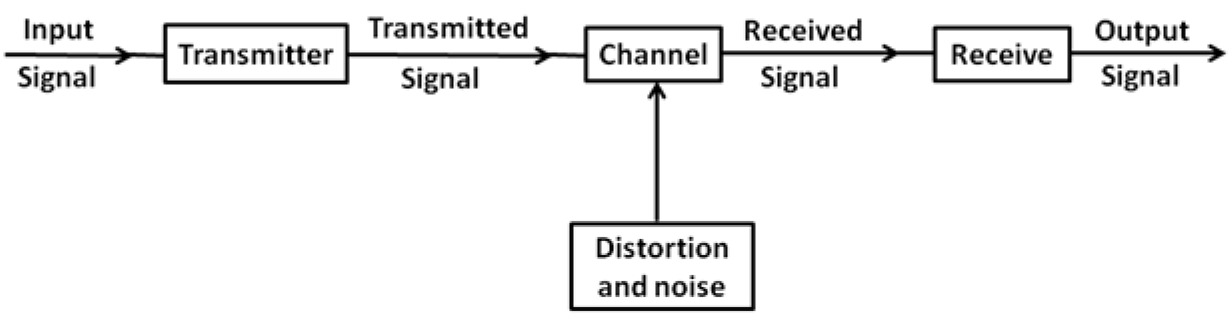
Property 1: The envelope of a narrow-band Gaussian process follows a Rayleigh distribution and the phase follows a uniform distribution in $(\mathbf{0}, \mathbf{2}\boldsymbol{\pi})$.

We note that

$$\sqrt{X_c{}^2(t) + X_s{}^2(t)} = R_X(t) \text{ and } \tan^{-1}\left\{\frac{X_s(t)}{X_c(t)}\right\} = \theta_X(t)$$

If X and Y are two independent $N(0, \sigma)$ then $R = \sqrt{X^2 + Y^2}$ follows a Rayleigh distribution and $\phi = tan^{-1}\frac{Y}{X}$ follows a uniform distribution in $(0,2\pi)$.

Noise in Communication Systems



In communication systems, the message to be transmitted to a far-off location is first converted into an electrical waveform called input signal, before being sent into the transmitter. The transmitter processes and modifies the input signal for efficient transmission. The transmitter output is then sent through the channel which is just a medium such as *wire*, *coaxial cable* or *optical fibre*. The channel output or the received signal is then reprocessed by the receiver which sends out the output signal. The output signal is converted to its original form, namely the message.

When the message is communicated in this manner, the signal is not only distorted by the channel but also contaminated along the path by undesirable signals that are generally referred to by the term **noise**. The noise can come from many external and internal sources and take many forms.

External noise includes interfering signals from nearby sources, man-made noise generated by faulty contact switches for the electrical equipment, by ignition radiation, fluorescent lights, natural noise from lighting and extraterrestrial radiation etc. Internal noise results from thermal motion of electrons in conductors, random emission and diffusion or recombination of charged carriers in electronic devices. By careful engineering techniques, the effects of many unwanted signals can be eliminated or minimized. But there always remain certain inescapable random signals that set a limit to system performance, $i.e.$, on the efficiency of communication.

One of the main reasons for introducing probability theory in the study of *Signal Analysis* is the random nature of noise. Because of this randomness, it is usual to describe noise as a random process and hence in terms of a probabilistic model. Such a model describes the noise amplitude or any other parameter by means of a probability density function $f(x)$ [x represents voltage]. For many important types of noise, the density function can be determined theoretically and for others it has been estimated empirically.

Certain properties of noise, such as mean value, mean square value and the root-mean square value can be found by using the probability density function.

However the probability density function does not describe a noise waveform sufficiently so as to determine its effect on the performance of a communication system. To achieve this, it is necessary to know how the noise changes with time. This information is provided by a mean-square voltage spectrum, called the **power spectrum** or **spectral density**,that represents the distribution of signal power as a function of frequency.

Thermal Noise

Thermal noise is the noise because of the random motion of free electrons in conducting media such as a resistor. Thermal noise generated in resistors and semiconductors is assumed to be a zero mean, stationary Gaussian random process $\{N(t)\}$ with a power spectral density that is flat over a very wide range of frequencies, $i.e.$, the graph of $S_{NN}(\omega)$ is a straight line parallel to the ω-axis. Since $S_{NN}(\omega)$ contains all frequencies in equal amount, the noise is also called **white**

Gaussian noise or **simply white noise** in analogy to white light which consists of all colours.

It is customary to denote the constant spectral density of white noise by $\frac{N_0}{2}\, or\, \frac{\eta}{2}$.

$i.e., S_{NN}(\omega) = \frac{N_0}{2}$

The autocorrelation function of the white noise is given by

$R_{NN}(\tau) = \frac{N_0}{2}\delta(\tau)$, since $\int_{-\infty}^{\infty} \frac{N_0}{2}\delta(\tau)e^{-i\omega\tau}d\tau = \frac{N_0}{2}$

The average power of the white noise $\{N(t)\}$ is given by

$$R_{NN}(0) = \int_{-\infty}^{\infty} S_{NN}(\omega)d\omega = \int_{-\infty}^{\infty} \frac{N_0}{2} d\omega \longrightarrow \infty$$

Therefore, the spectral density of $\{N(t)\}$ is not physically realisable. However, since the bandwidths of real processes are always finite and since

$$\int_{-\omega_B}^{\omega_B} S_{NN}(\omega)d\omega = N_0\omega_B < \infty$$

for any finite bandwidth, the spectral density $S_{NN}(\omega)$ can be used over finite bandwidths.

Band-limited white noise: Noise having a nonzero and constant spectral density over a finite frequency band and zero elsewhere is called **band-limited white noise.** $i.e.$, if $\{N(t)\}$ is a band-limited white noise then

$$S_{NN}(\omega) = \begin{cases} \frac{N_0}{2} & , \quad |\omega| \leq \omega_B \\ 0 & , \quad elsewhere \end{cases}$$

We give below a few properties of the band-limited white noise which can be easily verified by the reader.

1. $E\{N^2(t)\} = \frac{N_0\omega_B}{2\pi}$
2. $R_{NN}(\tau) = \frac{N_0\omega_B}{2\pi}\left(\frac{\sin \omega_B\tau}{\omega_B\tau}\right)$

3. $N(t)$ and $N\left(t+\frac{k\pi}{\omega_B}\right)$ are independent, where k is a nonzero integer.

Filters

Filtering is commonly used in electrical systems to reject undesirable signals and noise and to select the desired signal. A simple example of filtering occurs when we *tune* in a particular radio to *select* one of many signals.

Filtering actually means selecting carefully the transfer function $H(\omega)$ in a stable, linear, time-invariant system, so as to modify the spectral components of the input signal. The system function $H(\omega)$ or the linear system itself is referred to as filter, when it does the filtering.

The commonly used filters are narrow-band filters, $i.e.$, band pass and low pass filters.

If the system function $H(\omega)$ is defined as

$H(\omega) \neq 0$, for $\omega_0 - \frac{\varepsilon}{2} < \omega < \omega_0 + \frac{\varepsilon}{2}$ and $H(\omega) = 0$, otherwise

then the filter is called **a band pass filter**.

If $H(\omega) \neq 0$, for $-\frac{\varepsilon}{2} < \omega < \frac{\varepsilon}{2}$ and $H(\omega) = 0$, otherwise

then the filter is called **a low pass filter**.

The equation $S_{yy}(\omega) = S_{xx}(\omega)|H(\omega)|^2$ shows that the spectral properties of a signal can be modified by passing it through a linear time-invariant system with the appropriate transfer function. By carefully choosing $H(\omega)$, we can remove or filter out certain spectral components in the input. For example, let the input $X(t) = S(t) + N(t)$, where $S(t)$ is the signal of interest and $N(t)$ is an unwanted noise process. If the spectral densities of $\{S(t)\}$ and $\{N(t)\}$ are non-overlapping in the frequency domain, the noise $N(t)$ can be removed by passing $X(t)$ through a filter $H(\omega)$ that has a response of 1 for the range of frequencies occupied by the noise. But in most practical situations there is spectral overlap and the design of optimum filters to separate signal and noise is somewhat difficult. The discussion of this problem and the various optimum filters in common use such as matched filter and Wiener filter may be found in textbooks on Random Signal Analysis. It is beyond the scope of this syllabus.

Example 2: If $\{X(t)\}$ is a band limited process such that $S_{xx}(\omega) = 0$, when $|\omega| > \sigma$, prove that $2[R_{xx}(0) - R_{xx}(\tau)] \le \sigma^2\tau^2 R_{xx}(0)$.

Solution: $R_{xx}(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{xx}(\omega)e^{i\tau\omega}d\omega$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{xx}(\omega)\cos\tau\omega\, d\omega \qquad \text{(Since } s_{xx} \text{ is even)}$$

$$R_{xx}(0) - R_{xx}(\tau) = \frac{1}{2\pi}\int_{-\sigma}^{\sigma} S_{xx}(1-\cos\tau\omega)d\omega \qquad \text{(since} \{X(t)\} \text{ is band limited)}$$

$$= \frac{1}{2\pi}\int_{-\sigma}^{\sigma} S_{xx}(\omega)2\sin^2\left(\frac{\tau\omega}{2}\right)d\omega \qquad \text{... (1)}$$

From trigonometry, $|\sin\theta| \le \theta$

$\therefore\ sin^2\theta \le \theta^2$

$\therefore\ 2sin^2\left(\frac{\tau\omega}{2}\right) \le \frac{\tau^2\omega^2}{2}$ ... (2)

Inserting (2) in (1)

$$R_{xx}(0) - R_{xx}(\tau) \le \frac{1}{2\pi}\int_{-\sigma}^{\sigma} S_{xx}(\omega)\frac{\tau^2\omega^2}{2}d\omega$$

$$\le \frac{\sigma^2\tau^2}{4\pi}\int_{-\sigma}^{\sigma} S_{xx}(\omega)d\omega$$

$$\le \frac{\sigma^2\tau^2}{4\pi}\int_{-\infty}^{\infty} S_{xx}(\omega)d\omega$$

$i.e.,\quad R_{xx}(0) - R_{xx}(\tau) \le \frac{\sigma^2\tau^2}{2}R_{xx}(0)$

Example 3: Consider a white Gaussian noise of zero mean and power spectral density $\frac{N_0}{2}$ applied to a law pass RC filter whose transfer function is $H(f) = \frac{1}{1+i2\pi fRC}$. Find the autocorrelation function of the output random process.

Solution: The simple $RC -$ circuit for which the transfer function is given is a linear time – invariant system. The power spectral densities of the input $\{X(t)\}$ and the output $\{Y(t)\}$ of a linear system are connected by

$$S_{yy}(\omega) = S_{xx}(\omega)|H(\omega)|^2$$

In this problem the transfer function is expressed in terms of the frequency f.

Therefore, the above relation is

$$S_{yy}(f) = S_{xx}(f)|H(f)|^2$$

$$= \frac{1}{1+4\pi^2 f^2 R^2 C^2}\frac{N_0}{2} \qquad \text{(since the input is a white noise)}$$

$$\therefore \quad R_{yy}(\tau) = \frac{N_0}{2}\int_{-\infty}^{\infty}\frac{e^{i2\pi\tau f}}{1+4\pi^2 f^2 R^2 C^2}df$$

$$= \frac{N_0}{8\pi^2 R^2 C^2}\int_{-\infty}^{\infty}\frac{e^{i(2\pi\tau)f}\,df}{\left(\frac{1}{2\pi RC}\right)^2 + f^2} \qquad \dots (1)$$

Compare the integral in (1) with $\int_{-\infty}^{\infty}\frac{e^{imx}dx}{a^2+x^2} = \frac{\pi}{a}e^{-|m|\,a}$ ... (2)

Using (2) in (1)

$$R_{yy}(\tau) = \frac{N_0}{8\pi^2 R^2 C^2}\pi 2\pi RC\; e^{-|2\pi\tau|2\pi RC}$$

$$= \frac{N_0}{4RC}e^{-\frac{|\tau|}{RC}}$$

The mean square value of $\{Y(t)\}$ is given by $E\{Y^2(t)\} = R_{yy}(0) = \frac{N_0}{4RC}$

Example 4: If $Y(t) = A\cos(\omega_0 t + \theta) + N(t)$, where A is a constant, θ is a random variable with a uniform distribution in $(-\pi, \pi)$ and $\{N(t)\}$ is a band limited Gaussian white noise with a power spectral density.

$$S_{NN}(\omega) = \begin{cases} \frac{N_0}{2} \ , & for \ |\omega - \omega_0| < \omega_B \\ 0 \ , & elsewhere \end{cases}$$

find the power spectral density of $\{Y(t)\}$. Assume that $N(t)$ and θ are independent.

Solution:

$$Y(t_1)Y(t_2) = \{A\cos(\omega_0 t_1 + \theta) + N(t_1)\}\{A\cos(\omega_0 t_2 + \theta) + N(t_2)\}$$

$$= A^2\cos(\omega_0 t_1 + \theta)\cos(\omega_0 t_2 + \theta) + N(t_1)N(t_2)$$

$$+A\cos(\omega_0 t_1 + \theta)\,N(t_2) + A\cos(\omega_0 t_2 + \theta)\,N(t_1)$$

$$\therefore \quad R_{YY}(t_1, t_2) = E\{Y(t_1)Y(t_2)\}$$

$$= A^2 E\{\cos(\omega_0 t_1 + \theta)\cos(\omega_0 t_2 + \theta) + R_{NN}(t_1, t_2)\} +$$

$$AE\{\cos(\omega_0 t_1 + \theta)\}E\{N(t_2)\} + AE\{\cos(\omega_0 t_2 + \theta)\}\,E\{N(t_1)\}$$

(by independent)

$i.e.$, $R_{YY}(\tau) = \frac{A^2}{2}\cos\omega_0\tau + R_{NN}(\tau)$ [since $\{N(t)\}$ is stationary]

$$\therefore \quad S_{YY}(\omega) = \frac{A^2}{2}\int_{-\infty}^{\infty}\cos\omega_0\tau e^{-i\omega\tau}d\tau + S_{NN}(\omega)$$

$$= \frac{\pi A^2}{2}\{\delta(\omega - \omega_0) + \delta(\omega + \omega_0)\} + S_{NN}(\omega)$$

Where $S_{NN}(\omega)$ is given.

6.4

Markov chains

Markov Chain: If, for all n,

$P\{X_n = a_n | X_{n-1} = a_{n-1}, X_{n-2} = a_{n-2}, \dots, X_0 = a_0\} = P\{X_n = a_n | X_{n-1} = a_{n-1}\}$, then the process $\{X_n\}, n = 0,1, \dots, \dots$ is called a **Markov chain** and$(a_0, a_1, a_2, \dots, a_n, \dots)$ are called the **states** of the Markov chain. The conditional probability $P\{X_n = a_j | X_{n-1} = a_i\}$ is called the **one-step transition probability** from state a_i to state a_j at the n^{th} step (trial) and is denoted by $\boldsymbol{p_{ij}(n-1, n)}$. If the one-step transition probability does not depend on the step $i.e.$, $p_{ij}(n-1, n) = p_{ij}(m-1, m)$ the Markov chain is called a **homogeneous Markov chain** or the chain is said to have **stationary transition probabilities**. The use of the word *stationary* does not imply a stationary random sequence.

When the Markov chain is homogeneous, the one-step transition probability is denoted by p_{ij}. The matrix $P = \{p_{ij}\}$ is called (one-step) **transition probability matrix**,(t.p.m in short)

Note: The t.p.m of a Markov chain is a stochastic matrix, since $p_{ij} \geq 0$ and $\sum_j p_{ij} = 1$ for all i, $i.e.$, the sum of all the elements of any row of the t.p.m is 1. This is obvious because the transition from state a_i to any one of the states (including a_i itself) is a certain event.

The conditional probability that the process is in state a_j at step n, given that it was in state a_i at step 0, $i.e.$, $P\{X_n = a_j | X_0 = a_i\}$ is called the $\boldsymbol{n}$**-step transition probability** and denoted by $p_{ij}^{(n)}$.

Note: $p_{ij}{}^{(1)} = p_{ij}$.

Let us consider an example in which we explain how the t.p.m is formed for a Markov chain. Assume that a man is at an integral point of the x-axis between the origin and the point $x = 3$. He takes a unit step either to the right with probability

0.7 or to the left with probability 0.3, unless he is at the origin when he takes a step to the right to reach $x = 1$ or he is at the point $x = 3$, when he takes a step to the left to reach $x = 2$. The chain is called **Random walk with reflecting barriers**.

The t.p.m is given below:

$$
\begin{array}{cc}
 & \text{States of } X_n \\
 & \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\
\text{States of } X_{n-1} \quad \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \end{array} & \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0.3 & 0 & 0.7 & 0 \\ 0 & 0.3 & 0 & 0.7 \\ 0 & 0 & 1 & 0 \end{pmatrix}
\end{array}
$$

Note: p_{23} = the element in the 2nd row, 3rd column of this t.p.m is 0.7. This means that, if the process is at state 2 at step $(n-1)$, the probability that it moves to state 3 at step n is 0.7, where n is any positive integer.

Definition: If the probability that the process is in state a_i is $p_i (i = 1,2, \dots, k)$ at any arbitrary step, then the row vector $p = (p_1, p_2, \dots, p_k)$ is called the **probability distribution** of the process at that time. In particular, $p^{(0)} = \{p_1{}^{(0)}, p_2{}^{(0)}, \dots, p_k{}^{(0)}\}$ is the initial probability distribution.

Remark: The transition probability matrix together with the initial probability distribution completely specifies a Markov chain $\{\boldsymbol{X_n}\}$.

In the example given above, let us assume that the initial probability distribution of the chain is $p^{(0)} = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$. $i.e.$, $P\{X_0 = i\} = \frac{1}{4}$, $i = 0,1,2,3$.

Then we have, for the example given above,

$$P\{X_1 = 2|X_0 = 1\} = 0.7; \; P\{X_2 = 1|X_1 = 2\} = 0.3$$

$$P\{X_2 = 1, X_1 = 2|X_0 = 1\} = P\{X_2 = 1|X_1 = 2\}P\{X_1 = 2|X_0 = 1\}$$

$$= 0.3 \times 0.7 = 0.21 \qquad \dots (1)$$

$$P\{X_2 = 1, X_1 = 2, X_0 = 1\} = P\{X_0 = 1\}\, P\{X_2 = 1, X_1 = 2|X_0 = 1\}$$

$$= \frac{1}{4} \times 0.21 = 0.0525 \text{ [by (1)]} \qquad \dots (2)$$

$P\{X_3 = 3, X_2 = 1, X_1 = 2, X_0 = 1\}$

$$= P\{X_2 = 1, X_1 = 2, X_0 = 1\} \times P\{X_3 = 3 | X_2 = 1, X_1 = 2, X_0 = 1\}$$

$$= 0.0525 \times P\{X_3 = 3 | X_2 = 1\} \text{ (Markov property) [by (2)]}$$

$$= 0.0525 \times 0 = 0$$

Chapman-Kolmogorov Theorem:

If $\boldsymbol{P}$ is the t.p.m of a homogeneous Markov chain, then the $\boldsymbol{n}$-step t.p.m $\boldsymbol{P^{(n)}}$ is equal to $\boldsymbol{P^n}$. $\boldsymbol{i.e.}$, $\left[\boldsymbol{p_{ij}}^{(n)}\right] = \left[\boldsymbol{p_{ij}}\right]^{\boldsymbol{n}}$.

Proof: $p_{ij}^{(2)} = P\{X_2 = j | X_0 = i\}$, since the chain is homogeneous.

The state j can be reached from the state i in 2 steps through some intermediate state k.

Now $p_{ij}^{(2)} = P\{X_2 = j | X_0 = i\} = P\{X_2 = j, X_1 = k | X_0 = i\}$

$$= P\{X_2 = j | X_1 = k, X_0 = i\} P\{X_1 = k | X_0 = i\}$$

$$= p_{kj}^{(1)} p_{ik}^{(1)}$$

$$= p_{ik} p_{kj}$$

Since the transition from state i to state j in 2 steps can take place through any one of the intermediate states, k can assume the values 1,2,3, ... The transitions through various intermediate states are mutually exclusive.

Hence $p_{ij}^{(2)} = \sum_k p_{ik}\, p_{kj}$

$i.e.$, the ij-th element of 2 step t.p.m $=$ the ij-th element of the product of the 2 one-step t.p.m's

$i.e.$, $P^{(2)} = P^2$

Now $p_{ij}^{(3)} = P\{X_3 = j|X_0 = i\}$

$$= \sum_k P\{X_3 = j|X_2 = k\}\, P\{X_2 = k|X_0 = i\}$$

$$= \sum_k p_{kj} p_{ik}^{(2)}$$

$$= \sum_k p_{ik}^{(2)} p_{kj}$$

Similarly $p_{ij}^{(3)} = \sum_k p_{ik}\, p_{kj}^{(2)}$

i.e., $P^{(3)} = P^2 P = P P^2 = P^3$

Proceeding further in a similar way, we get

$$P^{(n)} = P^n$$

For example, consider the problem of Random walk with reflecting barriers discussed above, for which the t.p.m is

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0.3 & 0 & 0.7 & 0 \\ 0 & 0.3 & 0 & 0.7 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

$$P^2 = \begin{pmatrix} 0.3 & 0 & 0.7 & 0 \\ 0 & 0.51 & 0 & 0.49 \\ 0.09 & 0 & 0.91 & 0 \\ 0 & 0.3 & 0 & 0.7 \end{pmatrix}$$

From this matrix, we see that $P_{11}^{(2)} = 0.51$. This is so, because

$$P_{11}{}^{(2)} = \sum_{k=0}^{3} P_{1k}P_{k1} = P_{10}P_{01} + P_{11}P_{11} + P_{12}P_{21} + P_{13}P_{31}$$

$$= (0.3)(1) + (0)(0) + (0.7)(0.3) + (0)(0) = 0.51$$

Definition: A stochastic matrix P is said to be a **regular matrix**, if all the entries of P^m(for some positive integer m) are positive. A homogeneous Markov chain is said to be regular if its t.p.m is regular.

We state below two theorems without proof:

1. **If $\boldsymbol{p} = \{\boldsymbol{p_i}\}$ is the state probability distribution of the process at an arbitrary time, then that after one step is $\boldsymbol{pP}$, where $\boldsymbol{P}$ is the t.p.m of the chain and that after $\boldsymbol{n}$ steps in $\boldsymbol{pP^n}$.**
2. **If a homogeneous Markov chain is regular, then every sequence of state probability distributions approaches a unique fixed probability distribution called the *stationary (state) distribution* or *steady-state distribution* of the Markov chain.**
   That is, $\lim_{n \to \infty} \{\boldsymbol{p}^{(n)}\} = \boldsymbol{\pi}$, where the state probability distribution at step $\boldsymbol{n}$, $\boldsymbol{p}^{(n)} = \left(\boldsymbol{p_1}^{(n)}, \boldsymbol{p_2}^{(n)}, \dots, \boldsymbol{p_k}^{(n)}\right)$ $\boldsymbol{i.e.}, \boldsymbol{p}^{(n)} = \boldsymbol{p}^{(0)}\boldsymbol{P}^n$ and the stationary distribution $\boldsymbol{\pi} = (\boldsymbol{\pi_1}, \boldsymbol{\pi_2}, \dots, \boldsymbol{\pi_k})$ are row vectors.
3. Moreover, if P is the t.p.m of the regular chain, then $\pi P = \pi$(π is a row vector). Using this property of π, it can be found out, as in the worked examples given below:

Classification of States of a Markov Chain

If $P_{ij}{}^{(n)} > 0$ for some n and for all i and j, then every state can be reached from every other state. When this condition is satisfied, the Markov chain is said to be **irreducible**. The t.p.m of an irreducible chain is an irreducible matrix. Otherwise, the chain is said to be **nonirreducible or reducible.**

State i of a Markov chain is called a **return state**, if $P_{ii}{}^{(n)} > 0$ for some $n > 1$.

The period d_i of a return state i is defined as the greatest common divisor of all m such $p_{ii}{}^{(m)} > 0$, i.e., $d_i = GCD\{m: p_{ii}{}^{(m)} > 0\}$. State i is said to be **periodic with period d_i** if $d_i > 1$ and **aperiodic** if $d_i = 1$.

Obviously state i is aperiodic if $p_{ii} \neq 0$. The probability that the chain returns to state i, having started from state i, for the first time at the nth step (or after n transitions) is denoted by $f_{ii}{}^{(n)}$ and called the **first return time probability** or the **recurrence time probability**. $\{n, f_{ii}{}^{(n)}\}, n = 1,2,3, \ldots$, is the distribution of recurrence times of the state i.

If $F_{ii} = \sum_{n=1}^{\infty} f_{ii}{}^{(n)} = 1$, the return to state i is certain.

$\mu_{ii} = \sum_{n=1}^{\infty} n f_{ii}{}^{(n)}$ is called the **mean recurrence time** of the state i.

A state i is said to be persistent or recurrent if the return to state i is certain, i.e., if $F_{ii} = 1$. The state i is said to be transient if the return to state i is uncertain, i.e., if $F_{ii} < 1$. The state i is said to be nonnull persistent if its mean recurrence time μ_{ii} is finite and null persistent, if $\mu_{ii} = \infty$.

A nonnull persistent and aperiodic state is called ergodic.

We give below two theorems, without proof, which will be helpful to classify the states of a Markov chain.

1. If a Markov chain is irreducible, all its states are of the same type. They are all transient, all null persistent or all nonnull persistent. All its states are either aperiodic or periodic with the same period.
2. If a Markov chain is finite irreducible, all its states are nonnull persistent.

Example 1: The transition probability matrix of a Markov chain $\{X_n\}$, $n = 1, 2, 3, \ldots$, having 3 states 1, 2 and 3 is

$$\boldsymbol{P} = \begin{pmatrix} \mathbf{0.1} & \mathbf{0.5} & \mathbf{0.4} \\ \mathbf{0.6} & \mathbf{0.2} & \mathbf{0.2} \\ \mathbf{0.3} & \mathbf{0.4} & \mathbf{0.3} \end{pmatrix}$$

and the initial distribution is $\boldsymbol{p}^{(0)} = (\mathbf{0.7}, \mathbf{0.2}, \mathbf{0.1})$.

Find (i) $P\{X_2 = 3\}$ and (ii) $P\{X_3 = 2, X_2 = 3, X_1 = 3, X_0 = 2\}$.

Solution:

$$P^{(2)} = P^2 = \begin{pmatrix} 0.1 & 0.5 & 0.4 \\ 0.6 & 0.2 & 0.2 \\ 0.3 & 0.4 & 0.3 \end{pmatrix}\begin{pmatrix} 0.1 & 0.5 & 0.4 \\ 0.6 & 0.2 & 0.2 \\ 0.3 & 0.4 & 0.3 \end{pmatrix} = \begin{pmatrix} 0.43 & 0.31 & 0.26 \\ 0.24 & 0.42 & 0.34 \\ 0.36 & 0.35 & 0.29 \end{pmatrix}$$

(i) $P\{X_2 = 3\} = \sum_{i=1}^{3} P\{X_2 = 3 \mid X_0 = i\} P\{X_0 = i\}$

$$\begin{aligned} &= p_{13}^{(2)} P(X_0 = 1) + p_{23}^{(2)} P(X_0 = 2) + p_{33}^{(2)} P(X_0 = 3) \\ &= 0.26 \times 0.7 + 0.34 \times 0.2 + 0.29 \times 0.1 \\ &= 0.182 + 0.068 + 0.029 \\ &= 0.279 \end{aligned}$$

(ii) $P\{X_1 = 3 | X_0 = 2\} = p_{23} = 0.2$ ... (1)

$$P\{X_1 = 3, X_0 = 2\} = P\{X_1 = 3 | X_0 = 2\} \times P\{X_0 = 2\}$$

$= 0.2 \times 0.2 = 0.04$ [by (1)] ... (2)

$P\{X_2 = 3, X_1 = 3, X_0 = 2\}$

$= P\{X_2 = 3 | \, X_1 = 3, X_0 = 2\} \times P\{ X_1 = 3, X_0 = 2\}$

$= P\{X_2 = 3 | \, X_1 = 3\} \times P\{ X_1 = 3, X_0 = 2\}$ (by Markov property)

$= 0.3 \times 0.04$ [by (2)]

$= 0.012$ ... (3)

$P\{X_3 = 2, X_2 = 3, X_1 = 3, X_0 = 2\}$

$= P\{X_3 = 2|X_2 = 3, X_1 = 3, X_0 = 2\} \times P\{X_2 = 3, X_1 = 3, X_0 = 2\}$

$= P\{X_3 = 2|\ X_2 = 3\} \times P\{\ X_2 = 3, X_1 = 3, X_0 = 2\}$ (by Markov property)

$= 0.4 \times 0.012$ [by (3)]

$= 0.0048$

Example 2: A fair dice is tossed repeatedly. If X_n denotes the maximum of the numbers occurring in the first n tosses, find the transition probability matrix P of the Markov chain $\{X_n\}$. Find also P^2 and $P(X_2 = 6)$

Solution: State space: $\{1,2,3,4,5,6\}$

The t.p.m is formed using the following analysis.

Let $X_n =$ the maximum of the numbers occurring in the first n trials $=$ 3, say

Then $X_{n+1} = 3$, if the $(n+1)$th trial results in 1, 2 or 3

$= 4$, if the $(n+1)$th trial results in 4

$= 5$, if the $(n+1)$th trial results in 5

$= 6$, if the $(n+1)$th trial results in 6

$\therefore P\{X_{n+1} = 3|X_n = 3\} = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6} = \frac{1}{2}$

$P\{X_{n+1} = i|X_n = 3\} = \frac{1}{6}$, when $i = 4,5,6$

Therefore, the transition probability matrix of the chain is

$$P=\begin{pmatrix} \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ 0 & \frac{2}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ 0 & 0 & \frac{3}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ 0 & 0 & 0 & \frac{4}{6} & \frac{1}{6} & \frac{1}{6} \\ 0 & 0 & 0 & 0 & \frac{5}{6} & \frac{1}{6} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$P^2=\frac{1}{36}\begin{pmatrix} 1 & 3 & 5 & 7 & 9 & 11 \\ 0 & 4 & 5 & 7 & 9 & 11 \\ 0 & 0 & 9 & 7 & 9 & 11 \\ 0 & 0 & 0 & 16 & 9 & 11 \\ 0 & 0 & 0 & 0 & 25 & 11 \\ 0 & 0 & 0 & 0 & 0 & 36 \end{pmatrix}$$

Initial state probability distribution is $p^{(0)}=\left(\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6}\right)$ since all the values 1,2, … ,6 are equally likely.

$$\begin{aligned} P\{X_2=6\} &= \sum_{i=1}^{6} P\{X_2=6 \mid X_0=i\}\times P\{X_0=i\} \\ &= \frac{1}{6}\sum_{i=1}^{6} P_{i6}^{(2)} \\ &= \frac{1}{6}\times\frac{1}{36}(11+11+11+11+11+36) \\ &= \frac{91}{216} \end{aligned}$$

Example 3: A man either drives a car or catches a train to go to office each day. He never goes 2 days in a row by train but if he drives one day, then the next day he is just as likely to drive again as he is to travel by train. Now suppose that on the first day of the week, the man tossed a fair dice and drove to work if any only if a 6 appeared. Find (i) the probability that he takes a train on the third day and (ii) the probability that he drives to work in the long run.

Solution: The travel pattern is a Markov chain, with state space = (train, car)

The t.p.m of the chain is

$$P = \begin{array}{c c} & \begin{array}{cc} T & C \end{array} \\ \begin{array}{c} T \\ C \end{array} & \begin{pmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \end{array}$$

The initial state probability distribution is $p^{(1)} = \left(\frac{5}{6}, \frac{1}{6}\right)$,
since P(travelling by car) = P(getting 6 in the toss of the dice) = $\frac{1}{6}$

and P(travelling by train) = $\frac{5}{6}$

$$p^{(2)} = p^{(1)}P = \left(\frac{5}{6}, \frac{1}{6}\right)\begin{pmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} = \left(\frac{1}{12}, \frac{11}{12}\right)$$

$$p^{(3)} = p^{(2)}P = \left(\frac{1}{12}, \frac{11}{12}\right)\begin{pmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} = \left(\frac{11}{24}, \frac{13}{24}\right)$$

$\therefore$ P(the man travels by train on the third day) = $\frac{11}{24}$

Let $\pi = (\pi_1, \pi_2)$ be the limiting form of the state probability distribution or stationary state distribution of the Markov chain.

By the property of $\pi, \pi P = \pi$

$$i.e., (\pi_1, \pi_2)\begin{pmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} = (\pi_1, \pi_2)$$

$i.e., \frac{1}{2}\pi_2 = \pi_1$ ... (1)

and $\pi_1 + \frac{1}{2}\pi_2 = \pi_2 \Rightarrow \pi_1 = \frac{1}{2}\pi_2$ .... (2)

Equations (1) and (2) are one and the same.

Therefore, consider (1) or (2) with $\pi_1 + \pi_2 = 1$, since π is a probability distribution.

Solving, $\pi_1 = \frac{1}{3}$ and $\pi_2 = \frac{2}{3}$

$\therefore P\{\text{the man travels by car in the long run}\} = \frac{2}{3}$.

Example 4: Consider a communication system which transmits the 2 digits 0 and 1 through several stages. Let $X_n (n \geq 1)$ be the digit leaving the n^{th} stage of the system and X_0 be the digit entering the first stage (or leaving the 0^{th} stage). At each stage there is a constant probability q that the digit which enters will be transmitted unchanged (i.e., the digit will remain unchanged when it leaves) and the probability P otherwise ($i.e.$, the digit changes when it leaves), where $p + q = 1$. Write down the t.p.m P of the homogeneous two-state Markov chain $\{X_n\}$. Find P^m, P^∞ and the conditional probability that the digit entering the first stage is 0, given that the digit leaving the mth stage is 0. Assume that the initial state probability distribution is $p^{(0)} = (a, 1-a)$.

Solution: State space $= (0,1)$;

$$\begin{array}{cc} & \text{State of } X_{n+1} \\ & \begin{array}{cc} 0 & 1 \end{array} \\ P \equiv \text{State of } X_n \begin{array}{c} 0 \\ 1 \end{array} & \begin{pmatrix} q & p \\ p & q \end{pmatrix} \end{array}$$

Now $P^2 = \begin{pmatrix} q & p \\ p & q \end{pmatrix}\begin{pmatrix} q & p \\ p & q \end{pmatrix}$

$$= \begin{pmatrix} p^2 + q^2 & 2pq \\ 2pq & p^2 + q^2 \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{2}[(q+p)^2 + (q-p)^2] & \frac{1}{2}[(q+p)^2 - (q-p)^2] \\ \frac{1}{2}[(q+p)^2 - (q-p)^2] & \frac{1}{2}[(q+p)^2 + (q-p)^2] \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{2}+\frac{1}{2}r^2 & \frac{1}{2}-\frac{1}{2}r^2 \\ \frac{1}{2}-\frac{1}{2}r^2 & \frac{1}{2}+\frac{1}{2}r^2 \end{pmatrix}, \text{ where } q-p=r$$

$$P^3 = \begin{pmatrix} \frac{1}{2}+\frac{1}{2}r^3 & \frac{1}{2}-\frac{1}{2}r^3 \\ \frac{1}{2}-\frac{1}{2}r^3 & \frac{1}{2}+\frac{1}{2}r^3 \end{pmatrix}$$

The values of P^2 and P^3 make us guess that

$$P^m = \begin{pmatrix} \frac{1}{2}+\frac{1}{2}r^m & \frac{1}{2}-\frac{1}{2}r^m \\ \frac{1}{2}-\frac{1}{2}r^m & \frac{1}{2}+\frac{1}{2}r^m \end{pmatrix}$$

It is correct as can be proved by induction as follows:

$$P^{m+1} = \begin{pmatrix} q & p \\ p & q \end{pmatrix} \begin{pmatrix} \frac{1}{2}+\frac{1}{2}r^m & \frac{1}{2}-\frac{1}{2}r^m \\ \frac{1}{2}-\frac{1}{2}r^m & \frac{1}{2}+\frac{1}{2}r^m \end{pmatrix}$$

$$= \begin{bmatrix} \frac{q}{2}+\frac{q}{2}r^m+\frac{p}{2}-\frac{p}{2}r^m & \frac{q}{2}-\frac{q}{2}r^m+\frac{p}{2}+\frac{p}{2}+\frac{p}{2}r^m \\ \frac{p}{2}+\frac{p}{2}r^m+\frac{q}{2}-\frac{q}{2}r^m & \frac{p}{2}-\frac{p}{2}r^m+\frac{q}{2}+\frac{q}{2}r^m \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{2}+\frac{1}{2}r^{m+1} & \frac{1}{2}-\frac{1}{2}r^{m+1} \\ \frac{1}{2}-\frac{1}{2}r^{m+1} & \frac{1}{2}+\frac{1}{2}r^{m+1} \end{bmatrix}$$

$$\therefore \quad P^m = \begin{bmatrix} \frac{1}{2}+\frac{1}{2}r^m & \frac{1}{2}-\frac{1}{2}r^m \\ \frac{1}{2}-\frac{1}{2}r^m & \frac{1}{2}+\frac{1}{2}r^m \end{bmatrix}, \text{ where } m \text{ is a positive integer} \geq 1$$

$$P^{\infty} = \lim_{m\to\infty} (P^m) = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \text{since } |r| < 1$$

Now $P\{X_m = 0\,, X_0 = 0\} = P\{X_m = 0|X_0 = 0\} \times P\{X_0 = 0\} = aP_{00}^{(m)}$

and $P\{X_m = 0\,, X_0 = 1\} = bp_{10}^{(m)} b = 1 - a$

Now $P\{X_0 = 0|\, X_m = 0\} = \frac{p\{X_0=0\}\, P\{X_m=0|X_0=0\}}{P\{X_0=0\}\, P_{00}^{(m)} + p\{X_0=1\} P_{10}^{(m)}}$

$$= \frac{a\left\{\frac{1}{2}+\frac{1}{2}r^m\right\}}{a\left\{\frac{1}{2}+\frac{1}{2}r^m\right\}+b\left\{\frac{1}{2}-\frac{1}{2}r^m\right\}} = \frac{a(1+r^m)}{1+(a-b)r^m}, \text{ where } b = 1 - a$$