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 $\emptyset$.

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

(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 $A_1, A_2, \ldots, A_n$ 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 $A_1, A_2, \ldots, A_m$ 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

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

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 ${}^{n}C_{r} = \frac{{}^{n}P_{r}}{r!}$ 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 $n_{C_r}(r-1)! = \frac{\boldsymbol{n_{P_r}}}{\boldsymbol{r}}$

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 $= \frac{1}{2} \cdot \frac{\boldsymbol{n_{P_r}}}{\boldsymbol{r}}$.

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 $i = 1, 2, \dots, n$ such that $r_1 + r_2 + \cdots + r_\mathrm{n} = r$ is given by

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

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

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

(iii) atmost one object is permitted per box is $\boldsymbol{n}_{\boldsymbol{C}_{\boldsymbol{r}}}\ (\boldsymbol{r}\le\boldsymbol{n})$

(*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_{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_{P_3}

(iii) $- - -N$. No of words ending with N is 7_{P_3}

(iv) $E - -N$. No of words begin with E and end with N is 6_{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

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

1.1. Algebra of Sets and Counting Methods

Exercise:

1. The letters of the word MISSISSIPPI are arranged. Find
   a. All possible arrangements.
   b. All arrangements in which 4S's come together.
   c. All arrangements in which 4S's do not come together.
   d. All arrangements in which 4S's and 4I's come together.

2. If A and B stands in a line along with 10 other persons, then find the number of ways in which there are three persons between A and B.

3. If A and B stands in a circle along with 10 other persons, then find the number of ways in which there are three persons between A and B.

4. The letters of the word FLOWER are taken 4 at a time and arranged in all possible ways. Find the number of arrangements that
   a. Begins with F and ends with R .
   b. Contain the letter E.
5. Find the number of ways in which the letters of the word HOSTEL can be arranged so that
   a. The vowels may not be separated.
   b. The vowels occupy even places.

6. All the letters of the word EAMCET are arranged in all possible ways. Find the number of arrangements in which no two vowels are adjacent.

7. Find the number of arrangements that can be made by taking all the letters of the word ALGEBRA.

8. Find the number of arrangements that can be made by taking all the letters of the word MATHEMATICS such that
    a. 2M's come together.
    b. 2M's do not come together.

9. There are 5 maths, 6 physics and 8 chemistry books. How many ways are there to pick
    a. Two books not both on the same subject.
    b. Any two books.

10. How many ways are there to form a 3 letter words using the letters A, B, C, D, E, F.
    a. with repetition of letters.
    b. without repetition of any letter.
    c. without repetition that contain the letter E.
    d. with repetition that contain E.

11. Find the number of arrangements which can be made using all the letters of the word LAUGH if the vowels are adjacent.

12. If all permutations of the letters of the word AGAIN are arranged as in dictionary, find the 50^{th} word.

13. Find the number of ways in which any four letters can be (i) arranged and (ii) selected from the word CORGOO

14. Find the total number of (i) permutations and (ii) combinations of 4 letters that can be made out of the letters of the word EXAMINATION.

15. The digits 1,2,3,4 and 5 are given. Find

   a. 3 digit numbers without repetitions.
   b. 3 digit numbers with repetitions.
   c. 3 digit odd numbers without repetitions.
   d. 3 digit odd numbers with repetitions.
   e. 3 digit even numbers without repetitions.
   f. 3 digit even numbers with repetitions.
   g. 5 digit numbers without repetitions.
   h. 5 digit numbers with repetitions.

16. Find the number of 3 digit odd numbers that can be formed with digits 1,2,3,4, 5 when repetition of digits is

   a. Not allowed
   b. Allowed

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 $A_1, A_2, \ldots, A_n$ 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:

$$E_1 = \{HHH\};$$

$$E_2 = \{TTT\}$$

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

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

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

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

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

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

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

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

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

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

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

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.

1.2. Basic Concepts in Probability

Exercise:

1. A die is tossed twice and the number of dots facing up is counted and noted in the order of occurrence. Let us define

   A : Total number of dots showing is even

   B : Both dice are even

   C : Number of dots in dice differ by 1

   (i) Does A imply B or does B imply B?

   (ii) Find $\mathrm{A} \cap \mathrm{C}$.

2. A desk drawer contains five pens, three of which are dry.

   (i) The pens are selected at random one by one until a good pen found. The sequence of test results is noted. What is the sample space.

   (ii) Suppose that only the number and not the sequence, of pens tested in part(i) is noted. Specify the sample space.

3. Write the sample space corresponding to each of the following random experiment.

   (i) Select a ball from an urn containing balls numbered 1 to 50. Note the number of the ball.

   (ii) Select a ball from an urn containing balls numbered 1 to 4. Suppose that balls 1 and 2 are black and balls 3 and 4 are white. Note the number and colour of the ball you select.

   (iii) Toss a coin three times and note the sequence of heads and tails.

   (iv) Toss a coin four times and note the number of tails

   (v) Count the number of voice packets containing only silence produced from a group of N speakers in a 10-mins period.

   (vi) A block of information is transmitted repeatedly over a noisy channel until an error free block arrives at the receiver. Count the number of transmissions required.

   (vii) Pick a number at random between 0 and 1.

   (viii) Measure the time between two message arrivals at a message centre.

(ix) Measure the lifetime of a given computer memory chip in a specified environment.

(x) Pick two numbers at random between 0 and 1.

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

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

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 $A_1, A_2, \ldots, A_n, \ldots$ 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 $= {}^{11}C_2 = \frac{11 \times 10}{2} = 55$

(i) If both the drawn balls are white, they must be selected out of the 8 white balls and this can be done in ${}^{8}C_2 = \frac{8 \times 7}{2} = 28$ 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 ${}^{8}C_1 \times {}^{3}C_1 = 8 \times 3 = 24$

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

P1:

In a single throw with two uniform dice find the probability of throwing (i) Five, (ii) Eight

Solution:

Exhaustive number of cases in a single throw with two dice is $6^2 = 36$.

(i) Sum of '5' can be obtained on the two dice in the following mutually exclusive ways:
$(1,4), (2,3), (3,2), (4,1)$ $i.e.$, 4 cases in all, where the first and second number in the bracket ( ) refer to the numbers on the 1st and 2nd dice respectively.
$\therefore$ Required probability $= \frac{4}{36} = \frac{1}{9}$

(ii) The cases favourable to the event of getting sum of 8 on two dice are:
$(2,6), (3,5), (4,4), (5,3), (6,2)$ $i.e.$, 5 distinct cases in all.
$\therefore$ Required probability $= \frac{5}{36}$

P2:

If six dice are rolled, then find the probability that all show different faces

Solution:

In a random roll of six dice, the exhaustive number of cases is $n(S) = 6^6$.

Define the event E : All the six dice show different faces.

We can get any one of the six faces 1,2,3,4,5,6, on the first dice. For the happening of E, the second die must show any one of the remaining 5 faces, the third die must show any one of the remaining 4 faces, and so on, the 6^{th} die must show the remaining last face.

Hence, by the product rule, the number of cases favourable to the happening of E are $n(E) = 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 6!$.

$$\therefore P(E) = \frac{n(E)}{n(S)} = \frac{6!}{6^6}$$

P3:

The letters of the word 'failure' are arranged at random. Find the probability that the consonants may occupy only odd positions.

Solution:

There are 7 distinct letters in the word 'failure' and they can be arranged among themselves in 7! ways, which gives the exhaustive number of cases.

In the word 'failure' there are 4 vowels viz., a, i, u and e, and 3 consonants viz., f, l, r. These 3 consonants are to be placed in the 4 odd places viz., 1st, 3rd, 5th and 7th and this can be done in ${}^4C_3$ ways. Further these 3 consonants can be arranged among themselves in $3!$ ways and the remaining 4 vowels can be arranged among themselves in $4!$ ways. Associating all these operations, total number of favourable cases for the consonants to occupy only odd positions is ${}^4C_3 \times 3! \times 4!$.

$\therefore$ Required probability $= \dfrac{{}^4C_3 \times 3! \times 4!}{7!} = \dfrac{4 \times 3!}{7 \times 6 \times 5} = \dfrac{4}{35}$.

P4:

Find the probability that in a random arrangement of the letters of the word ASSASSINATION, the four S's, come consecutively.

Solution:

The word 'ASSASSINATION' contains 13 letters in which A occurs 3 times, S 4 times, N 2 times and I 2 times, and T and O once each.

Hence, the total number of permutations (exhaustive number of cases) is: $\frac{13!}{3!4!2!2!}$

The four S's can come consecutively if they occupy the 10 positions as given below.

(i) First four positions,
(ii) Second to 5th positions,
(iii) Third to 6th positions.
... ...
(x) Last four positions i.e., 10th to 13th positions.

In each of the above 10 arrangements, the remaining nine letters viz., 'AAINATION' of which 3 are A's, 2 are N's and 2 are I's, and the rest all different which can be arranged among themselves in $\frac{9!}{3!2!2!}$ ways.

Hence, total number of favourable cases is $\frac{10\times 9!}{3!2!2!}$

$\therefore$ Required probability $= \frac{10\times 9!}{3!2!2!} \div \frac{13!}{3!4!2!2!} = \frac{10\times 9!4!}{13!} = \frac{10\times 4!}{13\times 12\times 11\times 10} = \frac{2}{143}$

P5:

In a random arrangement of the letters of the word 'MATHEMATICS', find the probability that all the vowels come together.

Solution:

The total number of permutations of the letters of the word 'MATHEMATICS' are $\frac{11!}{2!2!2!}$, because it contains 11 letters, of which 2 are A's, 2 M's, 2 T's, and remaining are all different.

The word MATHEMATICS contains 4 vowels viz., AEAI, (2 A's being identical). To obtain the total number of arrangements in which these 4 vowels come together, we regard them as tied together, forming only one letter so that, the total number of letters in MATHEMATICS may be taken as $11 - 3 = 8$, out of which 2 are M's, 2 are T's and rest distinct and therefore, their number of arrangements is given by $\frac{8!}{2!2!}$

Further, the four vowels AEAI, two of which are identical and rest distinct can be arranged among themselves in $\frac{4!}{2!}$ ways. Hence, the total number of arrangements favourable to getting all vowels together is: $\frac{8!}{2!2!} \times \frac{4!}{2!}$

$\therefore$ Required probability $= \frac{8!4!}{2!2!2!} \div \frac{11!}{2!2!2!} = \frac{8!4!}{11!} = \frac{4!}{11\times10\times9} = \frac{4}{165}$

1.3. Definitions of Probability

Exercise:

1. What is the probability that a leap year selected at random will contain (a) 53 Tuesdays and (b) 53 Sundays or 53 Mondays?

2. In a single throw of two dice, what is the probability of getting (a) a total of 8 ; and (b) total different from 8 :

3. Prove that in a single throw with a pair of dice, the probability of getting the sum of 7 is equal to $\frac{1}{6}$ and the probability of getting the sum of 10 is equal to $\frac{1}{12}$

4. In the play of two dice, the thrower loses if his first throw is 2,4, or 12. He wins if his first throw is a 5 or 11. Find the ratio between his probability of losing and probability of winning in the first throw.
   Hint: Number of favourable cases for getting
   (a) 2,4 or 12 is $1 + 3 + 1 = 5$; (b) 5 or 11 is $4 + 2 = 6$

5. If a pair of dice is thrown, find the probability that the sum of the digits on them is neither 7 nor 11.

6. Tickets are numbered from 1 to 100. They are well shuffled and a ticket is drawn at random. What is the probability that the drawn ticket has :
   (a) an even number? (b) a number 5 or a multiple of 5?
   (c) a number which is greater than 75? (d) a number which is a square?

7. There are 17 balls, numbered from 1 to 17 in a bag. If a person selects one ball at random, what is the probability that the number printed on the ball will be an even number greater than 9?

8. An integer is chosen at random from the first 200 positive integers. What is the probability that integer chosen is divisible by 6 or 8?

9. One ticket is drawn at random from a bag containing 30 tickets numbered from 1 to 30. Find the probability that

   (a) It is multiple of 5 or 7; (b) It is multiple of 3 or 5

10. A number is chosen from each of the two sets :

$$A = \{1, 2, 3, 4, 5, 6, 7, 8, 9\} \;\; ; \quad B = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$$

If P_1 is the probability that the sum of the two numbers be 10 and P_2 the probability that their sum be 8, find $P_1 + P_2$.

11. A bag contains 7 white and 9 black balls. Two balls are drawn in succession at random. What is the probability that one of them is white and the other is black?

12. A bag contains eight balls, five being red and three white. If a man selects two balls at random from the bag. What is the probability that he will get one ball of each colour?

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

14. A bag contains 4 white, 5 red and 6 green balls. Three balls are drawn at random. What is the probability that a white, a red and a green ball are drawn?

15. A bag contains 8 black, 3 red and 9 white balls. If 3 balls are drawn at random, find the probability that
(a) all are black, (b) 2 are black and 1 is white, (c) 1 is of each colour,
(d) the balls are drawn in the order black, red and white, (e) None is red.

16. The Federal Match Company has forty female employees and sixty male employees. If two employees are selected at random, what is the probability that (i) both will be males, (ii) both will be females, (iii) there will be one of each sex?

17. If a single draw is made from a pack of 52 cards, what is the probability of securing either an ace of spades or a jack of clubs.

18. Four cards are drawn from a full pack of cards. Find the probability that two are spades and two are hearts?

19. What is the probability of getting 9 cards of the same suit in one hand at a game of bridge?

20. The letters of the word **Triangle** are arranged at random. Find the probability that the word so formed (a) starts with T, (b) ends with E, (c) starts with T and ends with E.

21. In a random arrangement of the letters of the word **VIOLENT**, find the chance that the vowels I,O,E occupy odd positions only.

22. In a random arrangement of the letters of the word **Allahabad**, find the chance that the vowels occupy the even places.

23. The letters of the word **ARRANGE** are arranged at random. Find the chance that : (a) The two R's come together (b) The two R's does not come together (c) The two R's and the two A's come together

24. (a) If the letters of the word REGUALTIONS be arrange at random, what is the chance that there will be exactly 4 letters between the R and the E?

(b) What is the probability that four S's come consecutively in the word

MISSISSIPPI?

25. A and B stand in a ring with 10 others persons. If the arrangement of the persons is at random, find the chance that (a) there are exactly three persons between A and B (b) A and B stand together

26. The first 12 letters of the English alphabet are written at random. What is the probability that (a) there are 4 letters between A and B (b) A and B are written down side by side.

27. Seven persons sit in a row at random. Find the chance that (a) three persons A, B, C sit together in a particular order (b) A, B, C sit together in any order (c) B and C occupy the end seats (iv) C always occupies the middle seat.

28. A six figure number is formed by the digits 4, 5, 6, 7, 8, 9 ; no digit being repeated. Find the probability that the number formed is (a) divisible by 5 (b) not divisible by 5.

29. Five digited numbers are formed from the digits 1, 2, 3, 4, 5. Find the chance that the number formed is greater than 2300.

Answers:

1. (a) $\frac{2}{7}$ (b) $\frac{3}{7}$
2. (a) $\frac{5}{36}$ (b) $\frac{31}{36}$
3.
4.
5. $\frac{7}{9} = 0.78$

6. (a) 0.5 (b) 0.2 (c) 0.25 (d) 0.10

7. $\frac{4}{17}$

8. $\frac{1}{4}$

9. (a) $\frac{1}{3}$ (b) $\frac{7}{15}$

10. $\frac{16}{81}$

11. $\frac{21}{40}$

12. $\frac{{}^5C_1 \times {}^3C_1}{{}^8C_2} = \frac{15}{28}$
13. $\frac{3}{28}$
14. $\frac{24}{91}$

15. (a) $\frac{14}{285}$ (b) $\frac{21}{95}$ (c) $\frac{18}{95}$ (d) $\frac{3}{95}$ (e) $\frac{34}{57}$

16. (a) 0.357, (b) 0.157, (c) 0.4848

17. $\frac{1}{26}$

18. $\frac{{}^{13}C_2 \times {}^{13}C_2}{{}^{52}C_4} = \frac{468}{20825}$

19. $4 \times {}^{13}C_9 \times {}^{39}C_4 \,/\, {}^{52}C_4$

20. (a) $\frac{1}{8}$ (b) $\frac{1}{8}$ (c) $\frac{1}{56}$

21. $\frac{{}^{4}C_3 \times 3!4!}{7!} = \frac{4}{35}$

22. $\frac{1\times5!}{2!} \div \frac{9!}{4!2!} = \frac{1}{126}$

23. (a) $\frac{6!}{2!} \div \frac{7!}{2!2!} = \frac{2}{7}$ (b) $(1260 - 360) \div 1260 = \frac{5}{7}$ (c)$\frac{5!}{1260} = \frac{2}{21}$

24. (a) $\frac{6}{55}$ (b) $\frac{4}{165}$

25. (a) $\frac{2}{11}$ (b) $\frac{2}{11}$

26.(a) $\frac{7}{66}$ (b) $\frac{1}{6}$

27. (a) $\frac{5!}{7!} = \frac{1}{42}$ (b) $\frac{5!\,3!}{7!} = \frac{1}{7}$ (c) $\frac{5\times2!}{7!} = \frac{1}{21}$ (d) $\frac{6!}{7!} = \frac{1}{7}$

28.(a) $\frac{1}{6}$ (b) $\frac{5}{6}$

29. $\frac{3!+3\times4!}{5!} = \frac{13}{20}$

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

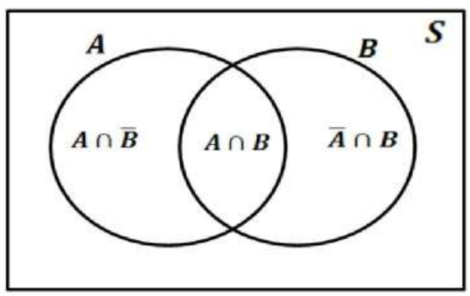
Proof: Since $\phi = \bar{S}$, $P(\phi) = P(\bar{S}) = 1 - P(S) = 1 - 1 = 0$ (by Axiom 2)

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

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,

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

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

(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

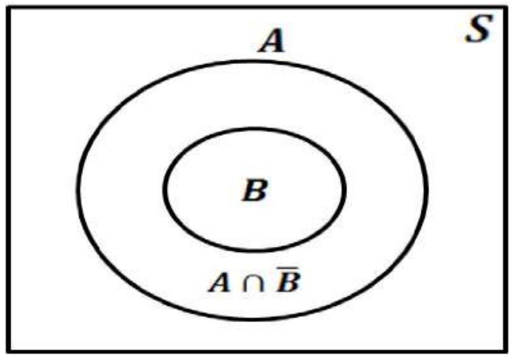
$$P(A) = P(A \cap B) + P(A \cap \bar{B})$$

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

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

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

Proof:



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

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

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

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

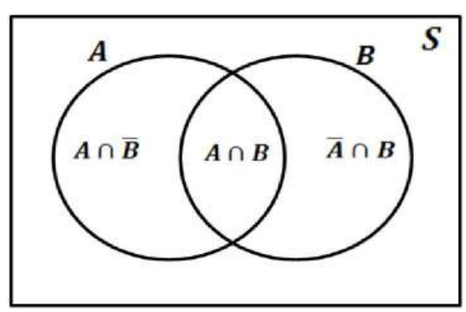
(ii) We have $P(A \cap \bar{B}) \ge 0$ (Axiom 1). Hence $P(A) - P(B) \ge 0 \Rightarrow P(B) \le P(A)$.
Thus, $B \subset A \Rightarrow P(B) \le P(A)$.

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

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

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 $A_1, A_2, A_3, \ldots, A_n$

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

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 $A_1, A_2, \dots, A_n$ are 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 A, B and C whose chances of solving it are $\frac{1}{3}$, $\frac{1}{4}$ and $\frac{1}{5}$ 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, $P(E_1 \cup E_2 \cup A_3) = 1 - P(\overline{E_1}) \cdot P(\overline{E_2}) \cdot P(\overline{E_3}) = 1 - \frac{2}{3} \times \frac{3}{4} \times \frac{4}{5} = 1 - \frac{2}{5} = \frac{3}{5}$

P1:

If one card is selected at random from a pack of 52 cards, find the probability that it is (i) a diamond or a spade (ii) an ace or a spade.

Solution:

(i) Let us denote the event of getting a diamond by A and a Spade by B. Then there are no common cards to diamonds and spades and hence $A \cap B = \phi$. Thus, A and B are mutually exclusive events. Further, $P(A) = \frac{13}{52} = \frac{1}{4}$ and $(B) = \frac{13}{52} = \frac{1}{4}$. Hence, by addition theorem, we have

P (a diamond or a spade) $= P(A \cup B) = P(A) + P(B) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$

(ii) Let us denote the event of getting an ace by A and a Spade by B. Then there is one card common for aces and spades and hence $A \cap B \neq \phi$. Thus, A and B are not mutually exclusive events. Further, $P(A) = \frac{4}{52}$, $P(B) = \frac{13}{52}$ and $P(A \cap B) = \frac{1}{52}$. Hence, by addition theorem, we have

P (an ace or a spade) $= P(A \cup B)$

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

$$= \frac{4}{52} + \frac{13}{52} - \frac{1}{52} = \frac{16}{52} = \frac{4}{13}.$$

P2:

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

Solution:

Let E, F and G denote the events that the adult reads newspapers A, B and C respectively. Then we are given:

$P(E) = 0.20$, $P(F) = 0.16$, $P(G) = 0.14$, $P(E \cap F) = 0.08$, $P(E \cap G) = 0.05$, $P(F \cap G) = 0.04$ and $P(E \cap F \cap G) = 0.02$. Thus,

$$P(E \cup F \cup G)$$

$$= P(E) + P(F) + P(G) - P(E \cap F) - P(E \cap G) - P(F \cap G) + P(E \cap F \cap G)$$

$$= 0.2 + 0.16 + 0.14 - 0.08 - 0.05 - 0.04 + 0.02 = 0.35$$

P3:

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

Proof: Given that A_1 and A_2 are independent events.

$P(A_1 \cap \overline{A}_2) = P(A_1) - P(A_1 \cap A_2)$ (By Theorem 3(ii))

$= P(A_1) - P(A_1) \cdot P(A_2)$ ($\because$ A_1 and A_2 are independent)

$= P(A_1)\big(1 - P(A_2)\big)$

$= P(A_1) \cdot P(\overline{A_2})$

$\Rightarrow P(A_1 \cap \overline{A_2}) = P(A_1) \cdot P(\overline{A_2})$

Thus, A_1 and $\overline{A_2}$ are independent.

P4:

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

Proof: Given that A_1 and A_2 are independent events.

Consider $P(\overline{A_1} \cap \overline{A_2}) = P(\overline{A_1 \cup A_2}) = 1 - P(A_1 \cup A_2)$

$$= 1 - \big(P(A_1) + P(A_2) - P(A_1 \cap A_2)\big)$$

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

$$= 1 - P(A_1) - P(A_2) + P(A_1) \cdot P(A_2) \text{ } (\because A_1 \& A_2 \text{ are independent})$$

$$= P(\overline{A_1}) - P(A_2)\big(1 - P(A_1)\big)$$

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

$$= P(\overline{A_1})\big(1 - P(A_2)\big)$$

$$= P(\overline{A_1}) \cdot P(\overline{A_2})$$

$\Rightarrow P(\overline{A_1} \cap \overline{A_2}) = P(\overline{A_1}) \cdot P(\overline{A_2})$

Thus, $\overline{A_1}$ and $\overline{A_2}$ are also independent.

1.4. Theorems in Probability

Exercises:

1. A card is drawn from a well shuffled pack of 52 cards. Find the probability that it is either a diamond or a king.
2. If $P(A) = 0.4$, $P(B) = 0.7$ and P (at least one of A and B)$= 0.8$, find P(only one of A and B).
3. Let A and B be the two possible outcomes of an experiment and suppose $P(A) = 0.4, P(A \cup B) = 0.7$ and $P(B) = p$.
   (i) For what choice of P are A and B mutually exclusive?
   (ii) For what choice of P are A and B independent?
4. An urn contains four tickets marked with numbers 112, 121, 211 and 222 and one ticket is drawn at random. Let $A_i(i = 1,2,3)$ be the event that i^{th} digit of the number of the ticket drawn is 1. Are A_1, A_2, A_3 (i) pairwise independent (ii) independent?
5. An engineer applies for a job in two firms X and Y. He estimates that the probability of his being selected in firm X is 0.7 and being rejected at Y is 0.5 and the probability of at least one of his applications being rejected is 0.6. What is the probability that he will be selected in one of the firms?
6. Probability that a man will be alive 25 years hence is 0.3 and the probability that his wife will be alive 25 years hence is 0.4. Find the probability that 25 years hence
   (i) both will be alive
   (ii) one the man will be alive
   (iii) only the woman will be alive
   (iv) none will be alive
   (v) at least one of them will be alive
7. The probability that a contractor will get a plumbing contract is $\frac{2}{3}$ and the probability that he will not get an electric contract is $\frac{5}{9}$. If the probability of getting at least one contract is $\frac{4}{5}$, what is the probability that he will get both the contracts?

8. A problem in probability is given to two students X and Y. The odds in favour of X solving the problem are 6 to 9 and against Y solving the problem are 12 to 10. If both X and Y attempt, find the probability of the problem being solved.
9. A piece of equipment will function only when all the three components A, B and C are working. The probability of A failing during one year is 0.15, that of B failing is 0.05 and that of C failing is 0.10. What is the probability that the equipment will fail before the end of the year?
10. Find the probability of throwing 6 at least once in six throws with a single die.
11. The odds that A speaks the truth are 3 : 2 and the odds that B speaks the truth are 5 : 3. In what percentage of cases are they likely to contradict each other on an identical point?
12. Three groups of children contain respectively 3 girls and 1 boy; 2 girls and 2 boys; 1 girl and 3 boys. One child is selected at random from each group. Find the probability that the selected consist of 1 girl and 2 boys.
13. If $P(A) = \frac{1}{4}$, $P(B) = \frac{2}{5}$ and $P(A \cup B) = \frac{1}{2}$, then find
    (i) $P(A \cap \bar{B})$
    (ii) $P(\bar{A} \cap \bar{B})$
14. If A, B and C are mutually exclusive and exhaustive event such that $P(A) = \frac{1}{2}P(B)$ and $P(B) = \frac{2}{3}P(C)$, find $P(A), P(B)$ and $P(C)$.
15. If $P(A) = 0.3, P(B) = 0.2$ and $P(C) = 0.1$ and A, B, C are independent events, find the probability of occurrence of atleast one of the three events A, B and C.

Answers:

1. $\frac{4}{13}$
2. 0.5
3. (i) $p = 0.3$ (ii) $p = 0.5$
4. (i) yes (ii) no
5. 0.8
6. (i) 0.12 (ii) 0.18 (iii) 0.28 (iv) 0.42 (v) 0.58
7. $\frac{14}{45}$

8. $\frac{37}{55}$

9. 0.27325

10. $1-\left(\frac{5}{6}\right)^{6}$

11. $\frac{19}{40}$

12. $\frac{13}{32}$

13. (i) 0.1 (ii) 0.85

14. $\frac{1}{6}, \frac{1}{3}, \frac{1}{2}$

15. 0.496

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

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

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

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

Also we have

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

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

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

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,

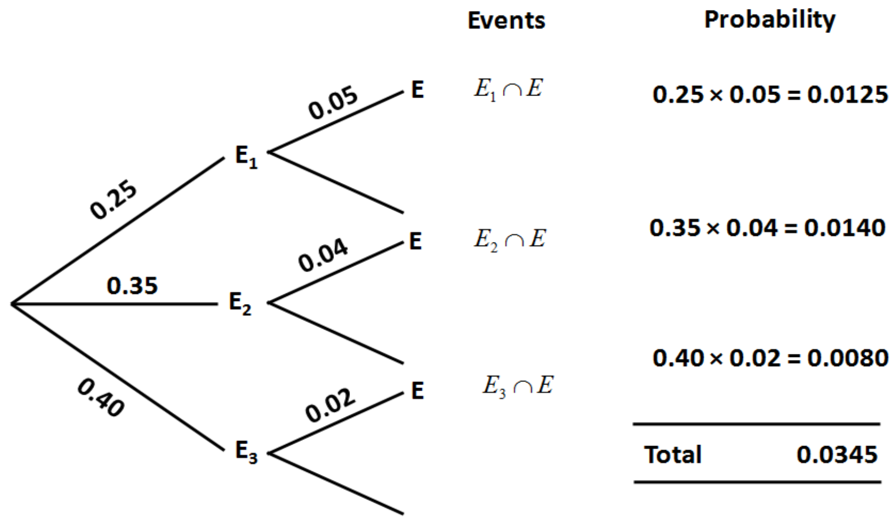
$$P(E_2|E) = \frac{0.0140}{0.0345} = \frac{28}{69} = 0.41\,; \qquad P(E_3|E) = \frac{0.0080}{0.0345} = \frac{16}{69} = 0.23$$

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:

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

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

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:

$$P(E_1|E) = \frac{P(E_1).P(E|E_1)}{P(E_1)P(E|E_1)+P(E_2)P(E|E_2)} = \frac{0.80\times0.30}{0.8\times0.30+0.2\times0.95} = \frac{0.240}{0.240+0.190} = \frac{24}{43}$$

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

(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 $\quad P(E_3|E) = 1 - \frac{33}{118} - \frac{55}{118} = \frac{30}{118}$)

P1:

A company has two plants to manufacture scooters. Plant I manufactures 80 percent of the scooters and plant II manufactures 20 percent. At plant I, 85 out of 100 scooters are rated standard quality or better. At plant II, only 65 out of 100 scooters are rated standard quality or better.

(i) What is the probability that scooter selected at random came from plant, I if it is known that the scooter is of standard quality?

(ii) What is the probability that the scooter came from plant II, if it is known that the scooter is of standard quality?

Solution:

Let us define the following events:

E_1: Scooter is manufactured by plant I

E_2: Scooter is manufactured by plant II

E: Scooter is rated as standard quality.

Then we are given:

$P(E_1) = 0.80$, $\quad P(E_2) = 0.20$; $\quad P(E|E_1) = 0.85 \quad P(E|E_2) = 0.65$

(i) Required probability is: (By Bayes' Rule)

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

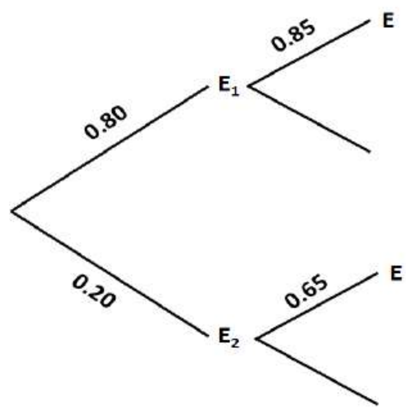
$$= \frac{0.80\times0.85}{0.80\times0.85\_0.20\times0.65} = \frac{0.68}{0.68+0.13} = \frac{0.68}{0.81} = 0.84$$

(ii) Required probability is given by:

$$P(E_2|E) = \frac{P(E_2)P(E|E_2)}{P(E_1)P(E|E_1)+P(E_2)P(E|E_2)} = \frac{0.20\times0.65}{0.80\times0.85+0.20\times0.65} = \frac{0.13}{0.81} = 0.16$$

(OR required probability is equal to$1 - 0.84 = 0.16$)

Aliter: **TREE DIAGRAM**



Events	Probability
$E \cap E_1$	$0.80 \times 0.85 = 0.68$
$E \cap E_2$	$0.20 \times 0.65 = 0.13$
Total	**0.81**

(i) $P(E_1|E) = \frac{0.68}{0.81} = 0.84$; (ii) $P(E_2|E) = \frac{0.13}{0.81} = 0.16$

P2:

In a class of 75 students, 15 were considered to be very intelligent, 45 as medium and the rest of them are below average. The probability that a very intelligent student fails in a viva – voice examination is 0.005; the medium student failing has a probability 0.05 ; and the corresponding probability for a below average student is 0.15. If a student is known to have passed the viva – voice examination, what is the probability that he is below average?

Solution:

Let us define the following events

E_1: The student of the class is very intelligent

E_2: The student is medium

E_3: The student is below average

A: The student passes in the viva- voice examination.

Then, we are given:

$P(E_1) = \frac{15}{75} = 0.2 \quad ; \quad P(E_2) = \frac{45}{75} = 0.6 \quad ; \quad P(E_3) = \frac{15}{75} = 0.2$

$P(A|E_1) = 1 - 0.005 = 0.995 \quad ; P(A|E_2) = 1 - 0.05 = 0.95$

$P(A|E_3) = 1 - 0.15 = 0.85$

Required to find $P(E_3|A)$.

If a student is known to have passed the viva – examination, the probability that he is below average, is given by Bayes' rule as follows.

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

$$= \frac{0.2 \times 0.850}{0.2 \times 0.995 + 0.6 \times 0.950 + 0.2 \times 0.850}$$

$$= \frac{0.170}{0.199 + 0.570 + 0.170} = \frac{0.170}{0.939} = 0.181$$

P3:

An urn contains four balls. Two balls are drawn at random and are found to be white. What is probability that all the balls are white?

Solution:

Since two balls are drawn and they are found to be white, the urn must contain at least two white balls. Let us define the following events:

E_i $(i = 2,\ 3, 4)$: The events that the urn contain i white balls.

E: The event that two white balls are drawn.

Since the events, E_2, E_3 and E_4 are equally likely, we have:

$P(E_2) = P(E_3) = P(E_4) = \frac{1}{3}$

$P(E|E_2) =$ Probability of drawing to white balls, given that the urn contains 2 white balls is $\frac{2_{C_2}}{4_{C_2}} = \frac{1}{6}$

Similarly, we have:

$P(E|E_3) = \frac{3_{C_2}}{4_{C_2}} = \frac{3}{6} = \frac{1}{2}$ and $P(E|E_4) = \frac{4_{C_2}}{4_{C_2}} = 1$

Required to find $P(E_4|E)$.

By Bayes' rule, we get:

$$P(E_4|E) = \frac{P(E_4)P(E|E_4)}{P(E_2)P(E|E_2)+P(E_3)P(E|E_3)+P(E_4)P(E|E_4)}$$

$$= \frac{\frac{1}{3}\times 1}{\frac{1}{3}\times\frac{1}{6}+\frac{1}{3}\times\frac{1}{2}+\frac{1}{3}\times 1} = \frac{\frac{1}{3}}{\frac{1}{18}+\frac{1}{6}+\frac{1}{3}} = \frac{\frac{1}{3}}{\frac{10}{18}} = \frac{3}{5} = 0.6$$

P4:

A speaks the truth 4 out of 5 times. A die is tossed. He reports that there is a six. What is the chance that actually there was six.

Solution:

Let us define the following events:

E_1: A Speaks truth ; E_2: A tells a lie ; E: A represents a six.

Form the given data, we have

$$P(E_1) = \frac{4}{5} \quad , \quad P(E_2) = \frac{1}{5}$$

$$P(E|E_1) = \frac{1}{6} \quad , \quad P(E|E_2) = \frac{5}{6}$$

By Baye's theorem, the required probability that actually there was six is $P(E_1/E)$ and

$$P(E_1/E) = \frac{P(E_1)P(E|E_1)}{P(E_1)P(E|E_1) + P(E_2)P(E|E_2)} = \frac{\frac{4}{5} \cdot \frac{1}{6}}{\frac{4}{5} \cdot \frac{1}{6} + \frac{1}{5} \cdot \frac{5}{6}} = \frac{4}{9}$$

1.5. Bayes' Theorem and Its Applications

Exercise

1. In 2004 there will be three candidates for the position of principal, in the college, Dr. Singhal, Mr. Mehra and Dr. Chatterji, whose chances of getting appointment are in the proportion $4 : 2 : 3$ respectvely. The probability that Dr. singhal if selected, will abolish co – education in the college is 0.3. The probability of Mr. Mehra and Dr. Chatterji doing the same are respectively 0.5 and 0.8. What is the probability that co – education will be abolished from the college in 2004?

2. (a) Suppose that one of three men, a politician, a businessman, and an educationist, will be appointed as the vice – chancellor of a university. The respective probabilities of their appointments are $0.50, 0.30, 0.20$. The probability that research activities will be promoted by these people if they are appointed are $0.30, 0.70$ and 0.80 respectively. What is the probability that research will be promoted by the new vice – chancellor?
(b) A manufacturing firm purchases a certain component for its manufacturing process from three sub – contractors A, B and C. These supply 60 percent, 30 percent and 10 percent of the firm's requirements, respective suppliers are defective items. On a particular day, a normal shipment arrives from each of the three suppliers and the contents get mixed. If a component is chosen at random from the day's shipment, what is the probability that it is defective?

3. Assume that a factory has two machines. Past records show that machine 1 produces 30% of the items of output and machine 2 were produces 70% of the items. Further, 5% of the items produced by machine 1 were defective and only 1% produced by machine 2 defective. If a defective item is drawn at random, what is the probability that it was produced by
(i) machine 1 , (ii) machine 2 ?

4. In a bolt factory machines A, B and C manufacture respectively $20\%, 30\%$ and 50% of the total of its output. Of them $5, 4$ and 2 percent respectively are defective bolts. A bolt is drawn at random from the product and it found to be defective. What is the probability that it was manufactured by machine B?

5. A factory produces a certain type of outputs by three types of machines. The respective daily production figures are:
Machine : $3{,}000$ Units ; Machine II : $2{,}500$ Units ; Machine III : $4{,}500$ Units
Past experience shows that 1 percent of the output produced by Machine I is defective. The corresponding fraction of defectives for the other two machines are 1.2 percent and 2 percent respectively. An item is drawn at random from the day's production run and is found to be defective, what is probability that it comes from the output of
(a) Machine 1, (ii) Machine II, (iii) Machine III?

6. Suppose that a product is produced in three factories A, B and C. It is known that factory A produces twice as many items as factory B, ad that factories B and C produces the same number of products. Assume that it is known that 2 oercent of the items produced by each of the factories A and B are defective while 4 percent of those manufactured by factory C are defective. All the items produced in three factories are stocked, and an item of product is selected at random. What is the probability that this item is defective?

7. A company has two plants to manufacture scooters. Plant I manufacutes 70% of the scooters and Plant II manufactures 30%. At Plant I, 80% of the scooters produced are of standard quality and at Plant II, 90% of the scooters produced are of standard quality. A scooter is picked at random and found to be of standard quality. What is the chance that it has come from Plant II?

8. Suppose there is a chance for a newly constructed building to collapse, whether the design is faulty or not. The chance that the design in faulty is 10%. The chance that the building collapses in 95%. If the design is faulty and otherwise it is 45%. It is seen that the building collapsed. What is the probability that it is due to faulty design.

Answers:

1. $\frac{23}{45}$
2. $a.\ 0.52$ b. 0.035
3. $(i)\ 0.682$, $(ii)\ 0.318.$
4. $\left(\frac{3}{8}\right) = 0.375$
5. a. $\frac{1}{5}$ b. $\frac{1}{5}$ c. $\frac{3}{5}$
6. 0.07
7. 0.03253
8. 0.19

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 $x_1, x_2, \dots, x_n, \dots$ .The function $p : \boldsymbol{R} \to \boldsymbol{R}$ defined as

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

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 $(x_i, p(x_i)), i = 1,2, \dots, n, \dots$ is called the **probability distribution** of the r.v.X.

In particular, if X takes the values $x_1, x_2, x_3, \dots, x_n$ 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 $b (a < b)$, $P(a \leq X \leq b) = \int_a^b f(x)dx.$

If the function $f(x)$ satisfies (i) $f(x) \geq 0 \quad \forall x \in \boldsymbol{R}$ 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

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

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

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

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)

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

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

P1:

Two dice are rolled at random. Obtain the probability distribution of the sum of the numbers on them.

Solution:

When two dice are rolled, the sample space S consists of $6^2 = 36$, sample points as shown.

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

Let X denote the sum of the numbers on the two dice. Then X is a random variable which can take the values $2, 3, 4, \dots, 12$ with the probability distribution given by:

Sum of numbers (x)	Favourable sample points	No. of favourable cases	$p(x)$
2	$(1,1)$	1	$\frac{1}{36}$
3	$(1,2),(2,1)$	2	$\frac{2}{36}$
4	$(1,3),(3,1),(2,2)$	3	$\frac{3}{36}$
5	$(1,4),(4,1),(2,3),(3,2)$	4	$\frac{4}{36}$
6	$(1,5),(5,1),(2,4),(4,2),(3,3)$	5	$\frac{5}{36}$
7	$(1,6),(6,1),(2,5),(5,2),(3,4),(4,3)$	6	$\frac{6}{36}$
8	$(2,6),(6,2),(3,5),(5,3),(4,4)$	5	$\frac{5}{36}$
9	$(3,6),(6,3),(4,5),(5,4)$	4	$\frac{4}{36}$
10	$(4,6),(6,4),(5,5)$	3	$\frac{3}{36}$
11	$(5,6),(6,5)$	2	$\frac{2}{36}$
12	$(6,6)$	1	$\frac{1}{36}$

Hence, the probability distribution of X is given by:

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

P2:

Four bad apples are mixed accidentally with 20 good apples. Obtain the probability distribution of the number of bad apples in a draw of 2 apples at random.

Solution:

Let X denote the number of bad apples drawn. Then X is a random variable which can take the values 0, 1 or 2. There are $4 + 20 = 24$ apples, in all and the exhaustive number of cases of drawing 2 apples is ${}^{24}C_2$.

$$\therefore P(X = 0) = \frac{{}^{20}C_2}{{}^{24}C_2} = \frac{20 \times 19}{24 \times 23} = \frac{95}{138}$$

$$P(X = 1) = \frac{{}^{4}C_1 \times {}^{20}C_1}{{}^{24}C_2} = \frac{2 \times 4 \times 20}{24 \times 23} = \frac{40}{138}$$

$$P(X = 2) = \frac{{}^{4}C_2}{{}^{24}C_2} = \frac{4 \times 3}{24 \times 23} = \frac{3}{138}.$$

Hence, the probability distribution of X is given by :

x :	0	1	2
$p(x)$:	$\frac{95}{138}$	$\frac{40}{138}$	$\frac{3}{138}$

P3:

The diameter of a telephone cable, say, x is assumed to be continuous random variable with p.d.f. $f(x) = kx(1 - x), 0 \leq x \leq 1$.

i. **Find k for which the above is a p.d.f.**
ii. **Determine b such that $P(x < b) = P(x > b)$.**

Solution:

i. $f(x) = kx(1 - x), 0 \leq x \leq 1$ is the p.d.f. of a continuous random variable x if $\int_0^1 f(x)dx = 1$. That is $k\int_0^1 x(1-x)dx = 1$

$$\Rightarrow k\left[\frac{x^2}{2} - \frac{x^3}{3}\right]_0^1 = 1 \Rightarrow k = 6$$

(i) Given $P(x < b) = P(x > b)$.That is,

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

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

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

$$\Rightarrow \left[3b^2 - 2b^3\right] = \left[1 - 3b^2 + 2b^3\right]$$

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

$$\Rightarrow (2b-1)\left(2b^2 - 2b - 1\right) = 0$$

$$\Rightarrow b = \frac{1}{2}, b = \frac{1 \pm \sqrt{3}}{2}$$

Hence, $b = \frac{1}{2}$ is the only value lying in $[0,1]$ and satisfying $P(x < b) = P(x > b)$

P4:

The time one has to spend in point of a bank counter is observed to be a random variable x with the p.d.f.

$$f(x)=\begin{cases} 0 & , \quad x<0 \\ \frac{1}{9}(x+1) & , \quad 0\le x<1 \\ \frac{4}{9}\left(x-\frac{1}{2}\right) & , \quad 1\le x<\frac{3}{2} \\ \frac{4}{9}\left(\frac{5}{2}-x\right) & , \quad \frac{3}{2}\le x<2 \\ \frac{1}{9}(4-x) & , \quad 2\le x<3 \\ \frac{1}{9} & , \quad 3\le x<6 \\ 0 & , \quad x\le 6 \end{cases}$$

Let A, B be the events defined as

A : **One waits between 0 and 2 minutes inclusive**

B : **One waits between 1 and 3 minutes inclusive**

i. **Show that $P(B|A)=\frac{2}{3}$**

ii. **Show that $P(\overline{A}\cap\overline{B})=\frac{1}{3}$**

Solution:

i.

$$P(A)=\int_0^2 f(x)dx=\int_0^1\frac{1}{9}(x+1)dx+\int_1^{\frac{3}{2}}\frac{4}{9}\left(x-\frac{1}{2}\right)dx+\int_{\frac{3}{2}}^{2}\frac{4}{9}\left(\frac{5}{2}-x\right)dx$$

$$=\frac{1}{9}\left[\frac{x^2}{2}+x\right]_0^1+\frac{4}{9}\left[\frac{x^2}{2}-\frac{x}{2}\right]_1^{\frac{3}{2}}+\frac{4}{9}\left[\frac{5x}{2}-\frac{x^2}{2}\right]_{\frac{3}{2}}^{2}$$

$$=\frac{1}{9}\left[\left(\frac{1}{2}+1\right)\right]+\frac{4}{9}\left[\left(\frac{9}{8}-\frac{3}{4}\right)\right]+\frac{4}{9}\left[(5-2)-\left(\frac{15}{4}-\frac{9}{8}\right)\right]=\frac{1}{2}$$

$$P(A\cap B)=P(1\le x\le 2)=\int_1^2 f(x)\,dx$$

$$=\int_1^{\frac{3}{2}}\frac{4}{9}\left(x-\frac{1}{2}\right)dx+\int_{\frac{3}{2}}^{2}\frac{4}{9}\left(\frac{5}{2}-x\right)dx$$

$$=\frac{4}{9}\left[\frac{x^2}{2}-\frac{x}{2}\right]_1^{\frac{3}{2}}+\frac{4}{9}\left[\frac{5x}{2}-\frac{x^2}{2}\right]_{\frac{3}{2}}^{2}$$

$$=\frac{4}{9}\left[\left(\frac{9}{8}-\frac{3}{4}\right)-\left(\frac{1}{2}-\frac{1}{2}\right)\right]+\frac{4}{9}\left[(5-2)-\left(\frac{15}{4}-\frac{9}{8}\right)\right]=\frac{1}{3}$$

$$P(B)=\int_1^3 f(x)\,dx$$

$$=\int_1^{\frac{3}{2}}\frac{4}{9}\left(x-\frac{1}{2}\right)dx+\int_{\frac{3}{2}}^{2}\frac{4}{9}\left(\frac{5}{2}-x\right)dx+\int_2^3\frac{1}{9}(4-x)\,dx$$

$$=\frac{4}{9}\left[\frac{x^2}{2}-\frac{x}{2}\right]_1^{\frac{3}{2}}+\frac{4}{9}\left[\frac{5x}{2}-\frac{x^2}{2}\right]_{\frac{3}{2}}^{2}+\frac{1}{9}\left[4x-\frac{x^2}{2}\right]_2^3$$

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

Thus $P(B|A)=\frac{P(A\cap B)}{P(A)}=\frac{\frac{1}{3}}{\frac{1}{2}}=\frac{2}{3}$

ii. $$P(\overline{A}\cap\overline{B})=P(\overline{A\cup B})=1-P(A\cup B)$$

$$=1-[P(A)+P(B)-P(A\cap B)]=1-\frac{2}{3}=\frac{1}{3}$$

Alternatively, $\overline{A}\cap\overline{B}$ means the waiting time more than 3 minutes.

So,

$$P(\overline{\mathrm{A}}\cap\overline{\mathrm{B}})=P(x>3)=\int_3^\infty f(x)\,dx=\int_3^6 f(x)\,dx+\int_3^\infty f(x)\,dx$$

$$=\frac{1}{9}\int_3^6 dx+0=\frac{1}{9}[6-3]=\frac{3}{9}=\frac{1}{3}$$

2.1. Random Variables

Exercise

1. State, with reasons, if the following probability distributions are admissible or not.

   (i)

x:	0	1	2
$p(x)$:	0.3	0.2	0.5

   (ii)

x:	– 1	0	2
$p(x)$:	0.4	0.4	0.3

   (iii)

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

   (iv)

x:	–2	–1	0	1	2
$p(x)$:	0.3	0.4	– 0.2	0.2	0.3

2. Two dice are thrown simultaneously and *getting a number less than 3* on a die is termed as a success. Obtain the probability distribution of the number of success.

3. Obtain the probability distribution of the number of sixes in two tosses of a die.
4. Obtain the probability distribution of number of heads of two tosses of a coin.

5. Three cards are drawn at random successively, with replacement, from a well shuffled pack of cards, getting 'a card of diamonds' is termed as a success. Obtain probability distribution of the number of successes.

6. Two cards are drawn without replacement, form a well shuffled pack of cards. Obtain the probability distribution of the number of face cards (Jack, Queen, King and Ace).

7. Five defective mangoes are accidently mixed with twenty good ones and by looking at them it is not possible to difference between them. Four mangoes are drawn at random from the lot. Find the probability distribution of x, the number of defective mangoes.

8. Two bad eggs are mixed accidentally with 10 good ones and three are drawn at random from the lot. Obtain the probability distribution of the number of bad eggs drawn.

9. An urn contains 6 red and 4 white balls. Three balls are known at random. Obtain the probability distribution of the number of white balls drawn.

10. Suppose that the life in hours of a certain part of radio tube is a continuous random variable x with p.d. f given by

$$f(x) = \begin{cases} \frac{100}{x^2}, & when\ x \geq 100 \\ 0\ , & otherwise \end{cases}$$

(i) What is the probability that all of three such tubes in a given radio set will have to be replaced during the first 150 hours of operation?

(ii) What is the probability that none of three of the original tubes will have to be replaced during that first 150 hours of operation?

(iii) What is the probability that a tube will last less than 200 hours if it is known that the tube still functioning after 150 hours of service.

(iv) What is the maximum number of tubes that many be inserted into a set so that there is a probability of 0.5 that after 150 hours of services all of them are still functioning?

Answers:

1.

(i) Yes

(ii) No, since $\sum p(x) > 1$

(iii) No, since $\sum p(x) < 1$

(iv) No, since $p(0) = -0.2$ which is not possible.

2.

x :	0	1	2
$p(x)$:	$\frac{4}{9}$	$\frac{4}{9}$	$\frac{4}{9}$

3.

x :	0	1	2
$p(x)$:	$\frac{25}{36}$	$\frac{10}{36}$	$\frac{1}{36}$

4.

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

5.

x:	0	1	2	3
$p(x)$:	$\frac{27}{64}$	$\frac{27}{64}$	$\frac{9}{64}$	$\frac{1}{64}$

6.

x:	0	1	2
$p(x)$:	$\frac{36_{C_2}}{52_{C_2}} = \frac{105}{221}$	$\frac{36_{C_1} \times 16_{C_1}}{52_{C_2}} = \frac{96}{221}$	$\frac{16_{C_2}}{52_{C_2}} = \frac{20}{221}$

7.

x:	0	1	2	3	4
$p(x)$:	$\frac{20_{C_4}}{25_{C_4}} = \frac{969}{2530}$	$\frac{5_{C_1} \times 20_{C_3}}{25_{C_4}} = \frac{1140}{2530}$	$\frac{5_{C_2} \times 20_{C_2}}{25_{C_4}} = \frac{380}{2530}$	$\frac{5_{C_3} \times 20_{C_1}}{25_{C_4}} = \frac{40}{2530}$	$\frac{5_{C_1}}{25_{C_4}} = \frac{1}{2530}$

8.

x:	0	1	2	3
$p(x)$:	$\frac{10_{C_3}}{12_{C_3}} = \frac{12}{22}$	$\frac{2_{C_1} \times 10_{C_2}}{12_{C_3}} = \frac{9}{22}$	$\frac{2_{C_2} \times 10_{C_1}}{12_{C_3}} = \frac{1}{22}$	0

9.

x:	0	1	2	3
$p(x)$:	$\frac{5}{30}$	$\frac{15}{30}$	$\frac{9}{30}$	$\frac{1}{30}$

10.

(i) $\frac{1}{3}$

(ii) $\frac{2}{3}$

(iii) $\frac{1}{4}$

(iv) 1.7 (= 2 approximately)

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

$$p(x) = P(X = x) = \begin{cases} \dfrac{1}{n} & , \quad x = 1,2, \dots , n \\ 0 & , \quad \text{otherwise} \end{cases}$$

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 $E\left(X^2\right) = \dfrac{1}{n}\sum_{i=1}^{n} i^2 = \dfrac{(n+1)(2n+1)}{6}$

Thus $V(X) = E\left(X^2\right) - \left(E(X)\right)^2 = \dfrac{(n+1)(n-1)}{12}$

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

$$F(x) = P(X \leq x) = \sum_{k=0}^{x} \binom{n}{k} p^k q^{n-k} \quad , x = 0,1,2, \dots, n$$

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: $\mathbf{A}$ and $\mathbf{B}$ play a game in which their chances of winning are in the ratio $\mathbf{3 : 2}$. 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 $\mathbf{\frac{4}{3}}$ 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

$$p(x) = P(X = x) = \begin{cases} \dfrac{e^{-\lambda}\lambda^x}{x!} & , \quad x = 0,1,2, \dots; \lambda > 0 \\ 0 & , \quad \text{otherwise} \end{cases}$$

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

$$V(X) = E(X^2) - \left(E(X)\right)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$$

$\Longrightarrow V(X) = \lambda$

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 $x = 0,1,2,\dots;\ \lambda > 0$ and

$P(Y = y) = \frac{e^{-\mu}\mu^y}{y!}$ for $y = 0,1,2,\dots;\ \mu > 0$.

Since $P(X = 1) = P(X = 2)$; $\lambda e^{-\lambda} = \frac{e^{-\lambda}\lambda^2}{2} \Longrightarrow \lambda^2 - 2\lambda = 0$

$\Longrightarrow \lambda(\lambda - 2) = 0 \Longrightarrow \lambda = 0, 2 \Longrightarrow \lambda = 2$ ($\because \lambda = 0$ 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

$$p(x) = P(X = x) = \begin{cases} \binom{x + r - 1}{r - 1} p^r q^x & , \quad x = 0,1,2, \dots \\ \qquad 0 & , \quad \text{otherwise} \end{cases}$$

Notation: $X \sim NB(r, p)$.

Note:

1. $\binom{x + r - 1}{r - 1} = \binom{x + r - 1}{x} \quad \left(\because \binom{n}{r} = \binom{n}{n - r}\right)$

$$= \frac{(x+r-1)(x+r-2)\dots(r+1)r}{x!}$$

$$= (-1)^r \frac{(-r)(-r-1)\dots(-r-x+2)(-r-x+1)}{x!} = (-1)^x \binom{-r}{x}$$

Thus, the p.m.f. of NBD can be written as

$$p(x) = \begin{cases} \binom{-r}{x} p^r (-q)^x & , \quad x = 0,1,2, \dots \\ \qquad 0 & , \quad \text{otherwise} \end{cases}$$

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 $\boldsymbol{P(X = x)}$: 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: $$\sum_{x=0}^{\min(n,M)} p(x) = \begin{cases} \sum_{x=0}^{n} \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} = \frac{\binom{N}{n}}{\binom{N}{n}} = 1 \; if \; \min(n,M) = n \\ \sum_{x=0}^{n} \frac{\binom{M}{x}\binom{N-M}{M-x}}{\binom{N}{M}} = \frac{\binom{N}{M}}{\binom{N}{M}} = 1 \; if \; \min(n,M) = M \end{cases}$$

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 $\boldsymbol{P_4}$**).

P1:

If $X \sim B(n, p)$, then show that $V(X) = npq$.

Proof: We already derived that $E(X) = \sum_{x=0}^{n} x\, p(x) = np.$

$$E(X^2) = \sum_{x=0}^{n} x^2 p(x) = \sum_{x=0}^{n} \left[x(x-1) + x\right] p(x)$$

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

$$= \sum_{x=2}^{n} x(x-1) \binom{n}{x} p^x q^{n-x} + np$$

$$= \sum_{x=2}^{n} x(x-1) \left(\frac{n}{x}\right)\left(\frac{n-1}{x-1}\right)\binom{n-2}{x-2} p^x q^{n-x} + np$$

$$= n(n-1)\, p^2 \sum_{x=2}^{n} \binom{n-2}{x-2} p^{x-2} q^{n-x} + np$$

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

$$\Rightarrow E(X^2) = n(n-1)\, p^2 + np \quad and \quad hence$$

$$V(X) = E(X^2) - (E(X))^2 = n(n-1)\, p^2 + np - n^2 p^2 = npq$$

P2:

If $\boldsymbol{\lambda = np}$, then $\boldsymbol{\lim_{n \to \infty} \lim_{p \to 0} \binom{n}{x} p^x q^{n-x} = \frac{e^{-\lambda}\lambda^x}{x!}}$.

Proof:

$$\binom{n}{x} p^x q^{n-x} = \binom{n}{x} p^x (1-p)^{n-x}$$

$$= \binom{n}{x} \left(\frac{p}{1-p}\right)^x (1-p)^n$$

$$= \frac{n(n-1)(n-2)\dots(n-x+1)}{x!} \frac{\left(\frac{\lambda}{n}\right)^x}{\left(1-\frac{\lambda}{n}\right)^x} \left(1-\frac{\lambda}{n}\right)^n$$

$$= \frac{n^x\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\dots\left(1-\frac{x-1}{n}\right)}{x!\left(1-\frac{\lambda}{n}\right)^x} \frac{\lambda^x}{n^x} \left(1-\frac{\lambda}{n}\right)^n$$

Then $\lim_{n \to \infty} \lim_{p \to 0} \binom{n}{x} p^x q^{n-x} = \frac{e^{-\lambda}\lambda^x}{x!} \quad \left(\because \lim_{n \to \infty} \left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda}\right)$

P3:

For negative binomial distribution, show that $V(X)=\frac{rq}{p^2}$.

Proof:

We already derived that $p(x)=\binom{-r}{x}p^r(-q)^x$

$$E(X)=\sum_{x=0}^{\infty}x\,p(x)=\frac{rq}{p}$$

$$E(X^2)=\sum_{x=0}^{\infty}x^2p(x)=\sum_{x=0}^{\infty}\left[x(x-1)+x\right]p(x)$$

$$=\sum_{x=2}^{\infty}x(x-1)\,p(x)+\sum_{x=0}^{\infty}x\,p(x)$$

$$=\sum_{x=2}^{\infty}x(x-1)\binom{-r}{x}p^r(-q)^x+\frac{rq}{p}$$

$$=\sum_{x=2}^{\infty}x(x-1)\left(\frac{-r}{x}\right)\left(\frac{-r-1}{x-1}\right)\binom{-r+2}{x-2}p^r(-q)^x+\frac{nq}{p}$$

$$=r(r+1)\,p^r(-q)^2\sum_{x=2}^{\infty}\binom{-r+2}{x-2}(-q)^{x-2}+\frac{nq}{p}$$

$$=r(r+1)\,p^rq^2(1-q)^{-(r+2)}+\frac{nq}{p}\quad\left(\because\sum_{x=2}^{\infty}\binom{-r+2}{x-2}(-q)^{x-2}=(1-q)^{-(r+2)}\right)$$

$$=r(r+1)\,p^rq^2p^{-(r+2)}+\frac{nq}{p}$$

$$\Rightarrow E(X^2)=r(r+1)\frac{q^2}{p^2}+\frac{nq}{p}$$

$$Thus,\ V(X) = E(X^2) - \left(E(X)\right)^2 = r(r+1)\frac{q^2}{p^2} + \frac{nq}{p} - \frac{r^2q^2}{p^2}$$

$$= \frac{r^2q^2}{p^2} + \frac{rq^2}{p^2} + \frac{nq}{p} - \frac{r^2q^2}{p^2}$$

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

$$\Rightarrow V(X) = \frac{rq}{p^2}$$

P4:

For hyper geometric distribution, show that mean and variance are given by

$$\boldsymbol{\mu = E(X) = \frac{nM}{n}} \quad \textbf{and}$$

$$\boldsymbol{\sigma^2 = V(X) = \frac{NM(N-M)(N-n)}{N^2(N-1)}}$$

when $\boldsymbol{\min(n, M) = n}$**.**

Proof:

We have $p(x) = \dfrac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$

$$E(X) = \sum_{x=0}^{n} x\, p(x) = \sum_{x=0}^{n} \frac{x\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

$$= \sum_{x=1}^{n} \frac{x\left(\frac{M}{x}\right)\binom{M-1}{x-1}\binom{N-M}{n-x}}{\binom{N}{n}}$$

$$= \frac{M}{\binom{N}{n}} \sum_{x=1}^{n} \binom{M-1}{x-1}\binom{N-M}{n-x} = \frac{M}{\binom{N}{n}}\binom{N-1}{n-1}$$

$$= \frac{M}{\left(\frac{N}{n}\right)\binom{N-1}{n-1}}\binom{N-1}{n-1} \quad \left(\because \sum_{x=1}^{n}\binom{M-1}{x-1}\binom{N-M}{n-x} = \binom{N-1}{n-1}\right)$$

$$\Rightarrow \mu = \frac{nM}{N}$$

$$Now, E\left(X^2\right) = \sum_{x=0}^{n} x^2 \ p\left(x\right) = \sum_{x=0}^{n} \left[x\left(x-1\right)+x\right] p\left(x\right) = \sum_{x=2}^{n} x\left(x-1\right) px + \sum_{x=0}^{n} xp\left(x\right)$$

$$= \sum_{x=2}^{n} \frac{x\left(x-1\right)\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} + \mu$$

$$= \sum_{x=2}^{n} \frac{x\left(x-1\right)\left(\frac{M}{x}\right)\left(\frac{M-1}{x-1}\right)\binom{M-2}{x-2}\binom{N-M}{n-x}}{\binom{N}{n}} + \mu$$

$$= \frac{M\left(M-1\right)}{\binom{N}{n}} \sum_{x=2}^{n} \binom{M-2}{x-2}\binom{N-M}{n-x} + \mu$$

$$= \frac{M\left(M-1\right)}{\binom{N}{n}} \binom{N-2}{n-2} + \mu \quad \left(\because \sum_{x=2}^{n} \binom{M-2}{x-2}\binom{N-M}{n-x} = \binom{N-2}{n-2}\right)$$

$$= \frac{M\left(M-1\right)}{\left(\frac{N}{n}\right)\left(\frac{N-1}{n-1}\right)\binom{N-2}{n-2}} \binom{N-2}{n-2} + \mu$$

$$\Rightarrow E\left(X^2\right) = \frac{n\left(n-1\right)M\left(M-1\right)}{N\left(N-1\right)} + \frac{nM}{N}$$

Thus, $V(X) = E(X^2) - (E(X))^2$

$$= \frac{n(n-1)M(M-1)}{N(N-1)} + \frac{nM}{N} - \frac{n^2M^2}{N^2}$$

$$= \frac{NM(N-M)(N-n)}{N^2(N-1)} \text{ (on simplification)}$$

2.4. Discrete Probability Distributions

Exercise

1. Ten coins are tossed simultaneously. Find the probability of getting at least seven heads.

2. A multiple choice test consists of 8 questions with 3 answers to each question (of which only one is correct). A student answers each question by rolling a balanced die and checking the first answer if he gets 1 or 2, the second answer if he gets 3 or 4 and the third answer if he gets 5 or 6. To get a distinction, the student must secure at least 75% correct answers. If there is no negative marking, what is the probability that the student secures a distinction?

3. In a precision bombing attack there is a 50% chance that any one bomb will strike the target. Two direct hits are required to destroy the target completely. How many bombs must be dropped to give a 99% chance or better of completely destroying the target?

4. A manufacturer of pins knows that 5% of his product is defective. If he sells pins in boxes of 100 and guarantees that not more than 10 pins will be defective, what is the probability that a box fails to meet the guaranteed quality?

5. An insurance company insures $4{,}000$ people against loss of both eyes in a car accident. Based on previous data, the rates were computed on the assumption that on the average 10 persons in $1{,}00{,}000$ will have car accident each year that result in this type of injury. What is the probability that more than 3 of the insured will collect on their policy in a given year?

6. A manufacturer, who produces medicine bottles, finds that 0.1% of the bottles are defective. The bottles are packed in boxes containing 500 bottles. A drug manufacturer buys 100 boxes from the product of bottles. Find how many boxes will contain (i) no defective, and (ii) at least two defectives.

7. Six coins are tossed 6,400 times. Using Poisson distribution, find the approximate probability of getting six heads two times.

8. Find the probability that a person tossing 3 coins will get either all heads or all tails, for the second time on the fifth toss.

9. If the probability is 0.4 that a child exposed to a certain contagious disease will catch it, what is the probability that the tenth child exposed to the disease will be third to catch it.

10. If the probabilities of having a male or female child are both 0.5, find the probability that
    a) a family's fourth child is their first son
    b) a family's seventh child is their second daughter
    c) a family's tenth child is their fourth or fifth son

Answers:

1. $\frac{176}{1024}$

2. $\frac{129}{729\times9}$

3. 11
4. $1 - e^{-5}\sum_{x=0}^{10}\frac{5^x}{x!}$

5. 0.0008

6. (i) 61 (ii) 9

7. $\frac{e^{-100}(100)^2}{2}$

8. $\binom{4}{1}\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)^3$

9. $\binom{9}{2}(0.4)^3(0.6)^7$

10. a) $(0.5)^4$ b) $6(0.5)^7$ c) $210(0.5)^{10}$

2.5

Continuous Probability Distributions

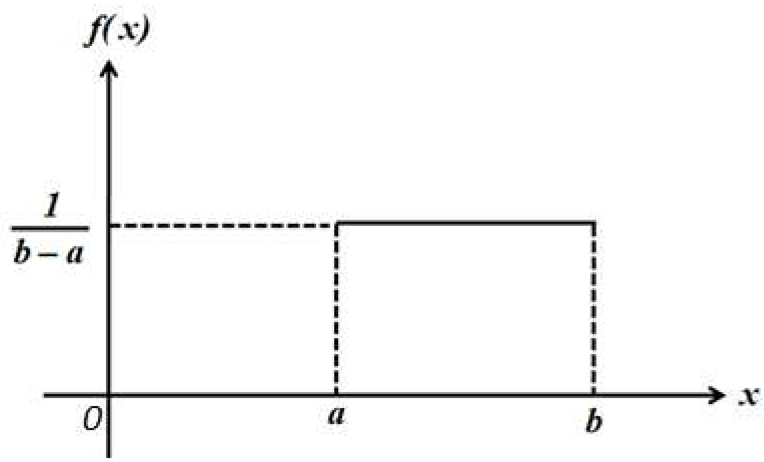
The **continuous probability distributions** are used in a number of applications in *engineering*. For example in *error analysis*, given a set of data or probability distribution, it is possible to estimate the probability that a measurement (temperature, pressure, flow rate) will fall within a desire 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

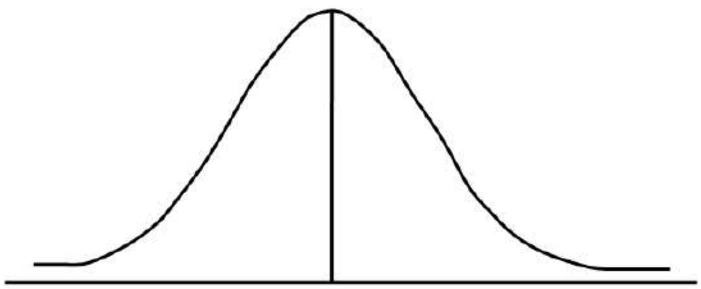
$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} exp\left\{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right\} dt$$

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

$$P(\mu - \sigma < X < \mu + \sigma) = P(-1 < Z < 1)$$

$$= \int_{-1}^{1} g(z)\,dz$$

$$= 2\int_0^{1} g(z)\,dz \qquad \text{(by symmetry)}$$

$$= 2 \times 0.3413 \qquad \text{(from table)}$$

$\Rightarrow P(\mu - \sigma < X < \mu + \sigma) = 0.6826$

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)

$\Rightarrow P(\mu - 2\sigma < X < \mu + 2\sigma) = 0.9544$

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)

$\Rightarrow P(\mu - 3\sigma < X < \mu + 3\sigma) = 0.9973$

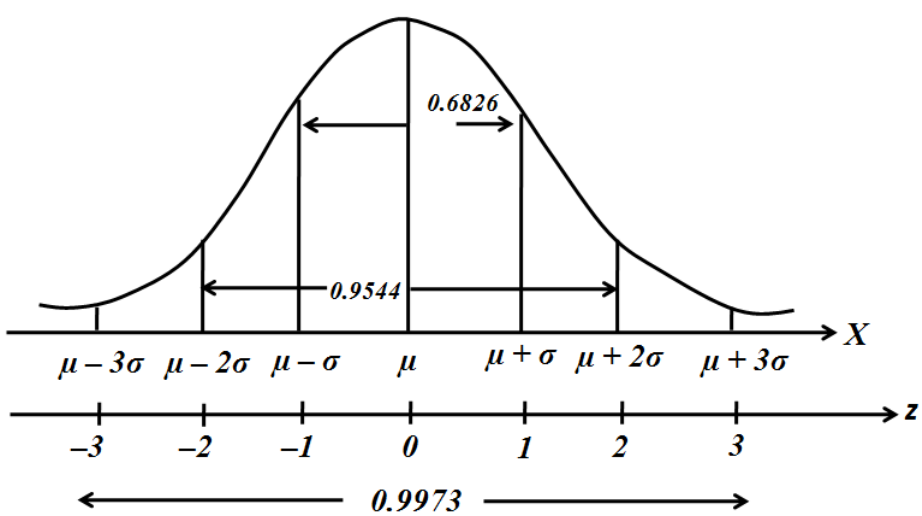
Thus, the probability that a normal variate X lies outside the range $\mu \pm 3\sigma$ is given by

$P(|x - \mu > 3\sigma|) = P(|Z| > 3) = 1 - P(|Z| \leq 3)$

$= 1 - P(-3 \leq Z \leq 3) = 1 - 0.9973 = 0.0027$

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

$$E(X)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}(\mu+\sigma z)e^{-\frac{1}{2}z^2}dz$$

$$=\mu.\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{1}{2}z^2}dz+\frac{\sigma}{\sqrt{2\pi}}\int_{-\infty}^{\infty}ze^{-\frac{1}{2}z^2}dz=\mu\times1+0$$

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

$$V(X) = E\big(X - E(X)\big)^2 = E(X-\mu)^2 \qquad (\because E(X)=\mu)$$

$$=\int_{-\infty}^{\infty}(x-\mu)^2 f(x)\,dx=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}(x-\mu)^2e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}dx$$

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:

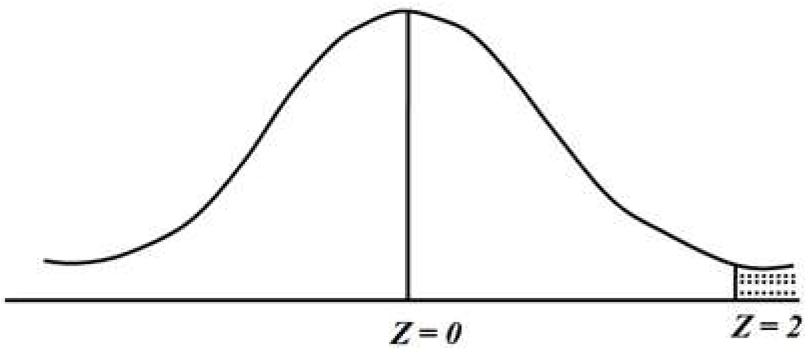
(a) it is given that $\mu = 12$ and $\sigma = 4$ $i.e., X \sim N(12, 16)$

(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) $P(0 \leq X \leq 12) = P\left(\frac{0-12}{4} \leq \frac{X-12}{4} \leq \frac{12-12}{4}\right)$

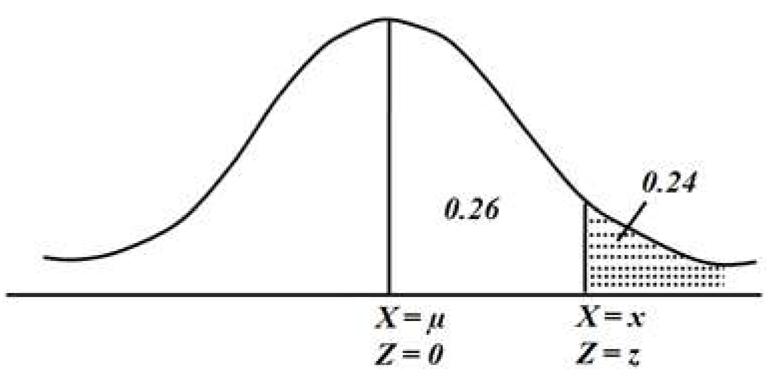
$= P(-3 \leq Z \leq 0)$

$= P(0 \leq Z \leq 3)$ (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 $z = \frac{x-12}{4}$



$\Longrightarrow P(0 < Z < z) = 0.5 - 0.24 - 0.26$

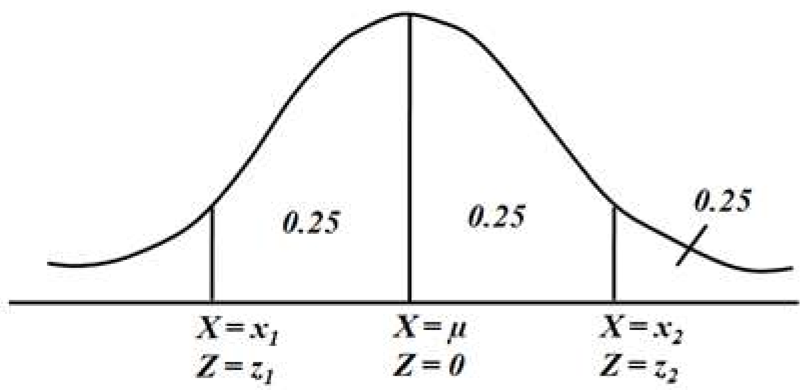
$\therefore$ From normal tables, corresponding to probability 0.26, value of $z = 0.71$(approximately)

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

$\frac{x_2-12}{4} = 0.67 \implies x_2 = 12 + 4 \times 0.67 = 14.68$

Similarly, $z_1 = -z_2 \implies \frac{x_2-12}{4} = -0.67 \Longrightarrow x_1 = 12 - 4 \times 0.67 = 9.32$

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 $Z = \frac{X-1000}{200} \sim N(0,1)$

$= P(Z > 1) = 0.5 - P(0 < Z < 1)$

$= 0.5 - 0.3413 = 0.1587$

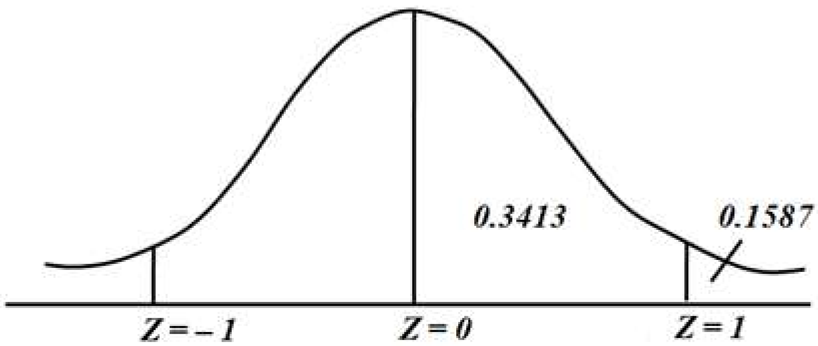
$\therefore$ Out of 10,000 bulbs, number of bulbs which fail in the first 800 hours is

$10{,}000 \times 0.1587 = 1{,}587.$

(ii) Find $P(800 < X < 1200) = P\left(\frac{800-1000}{200} < \frac{X-1000}{200} < \frac{1200-1000}{200}\right)$

$= P(-1 < Z < 1) = 2.P(0 < Z < 1)$

$= 2 \times 0.3413 = 0.6826$



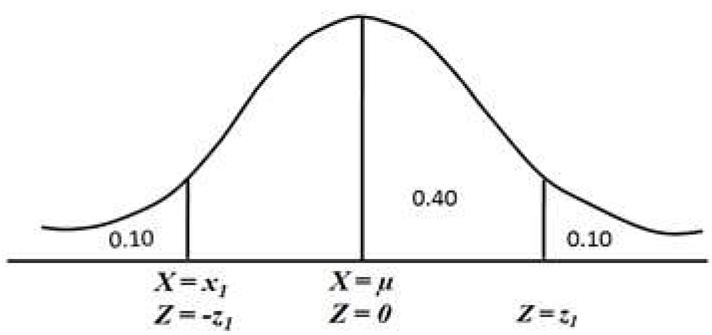
Hence, the expected number of bulbs with life between 800 and 1200 hours of burning life is $10{,}000 \times 0.6826 = 6{,}826$.

(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.

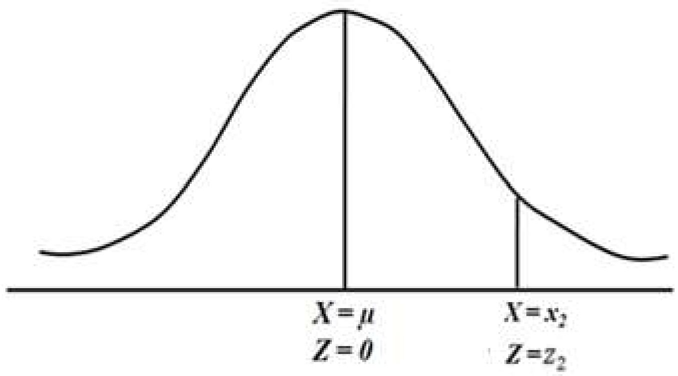
(b) Let 10% of the bulbs be still burning after x_2 hours of burning life. Then we have

$$P(X > x_2) = 0.10 \Rightarrow P\left(\frac{X-1000}{200} > \frac{x_2-1000}{200}\right) = 0.10$$

$\Rightarrow P(Z > z_2) = 0.10$,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,

$$\lim_{n\to\infty} P(k_1 \le X \le k_2) = \lim_{n\to\infty} P\left(\frac{k_1 - np}{\sqrt{npq}} \le Z \le \frac{k_2 - np}{\sqrt{npq}}\right) = \frac{1}{\sqrt{2\pi}} \int_{z_1}^{z_2} e^{-\frac{1}{2}z^2} dz$$

$$\text{where } z_1 = \frac{k_1 - np}{\sqrt{npq}} \text{ and } z_2 = \frac{k_2 - np}{\sqrt{npq}}$$

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

$$P(4 \le X \le 7) = \sum_{x=4}^{7} p(x) = \sum_{x=4}^{7} \binom{n}{x} \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{10-x} = \sum_{x=4}^{7} \binom{n}{x} \left(\frac{1}{2}\right)^{10}$$

$$= \frac{\binom{10}{4} + \binom{10}{5} + \binom{10}{6} + \binom{10}{7}}{1024} = \frac{792}{1024} = 0.7734$$

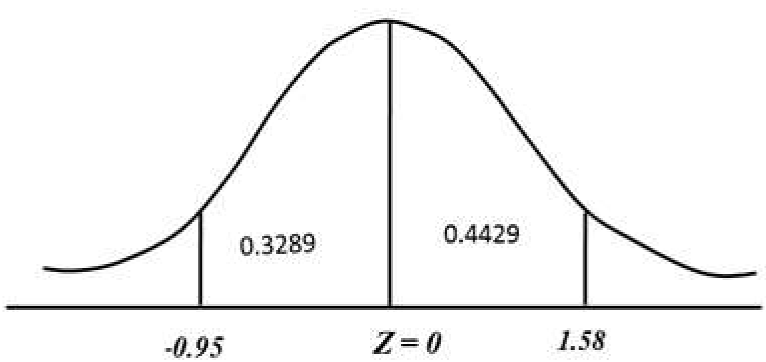
(b) The discrete binomial probability distribution is approximated to continuous normal probability distribution. The integers 4,5,6,7 lie in the interval (3.5 to 7.5). Thus,

$$P(4 \le X \le 7) = P(3.5 \le X \le 7.5) = P\left(\frac{3.5-5}{\sqrt{25}} \le Z \le \frac{7.5-5}{\sqrt{2.5}}\right)$$

$$= P(-0.95 \le Z \le 1.58) = P(-0.95 \le Z \le 0) + P(0 \le Z \le 1.58)$$

$$= P(0 \le Z \le 0.95) + P(0 \le Z \le 1.58)$$

$$= 0.3289 + 0.4429 = 0.7718$$



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.

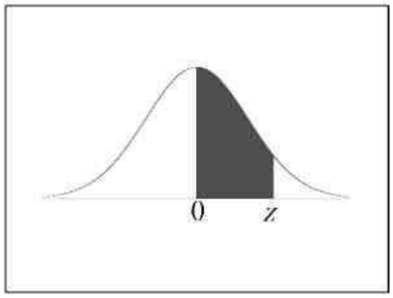
Let X denote the number of arrivals during an interval of t seconds.

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 $x = 0,1,2,\ldots;\ \ t \geq 0$, 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

P1:

The weights in pounds of parcels arriving at a package delivery company's warehouse can be modeled by an $N(5, 16)$ normal random variable X.

a) **What is the probability that a randomly selected parcel weighs between 1 and 10 pounds?**
b) **What is the probability that a randomly selected parcel weighs more than 9 pounds?**

Solution:

Since $X \sim N(5,16)$, we have $\mu = 5$ and $\sigma^2 = 16$.

a) $$\begin{aligned} P(1 < X < 10) &= P\left(\frac{1-5}{4} < Z < \frac{10-5}{4}\right) = P(-1 < Z < 1.25) \\ &= P(-1 < Z < 0) + P(0 < Z < 1.25) \\ &= P(0 < Z < 1) + P(0 < Z < 1.25) \\ &= 0.3413 + 0.3943 = 0.7356 \end{aligned}$$ (Use table)

b) $$\begin{aligned} P(X > 9) &= P\left(Z > \frac{9-5}{4}\right) = P(Z > 1) \\ &= 0.5 - P(0 < Z < 1) \\ &= 0.5 - 0.3413 \\ &= 0.1587 \end{aligned}$$ (Use table)

P2:

The marks obtained by a number of students for a certain subject are assumed to be approximately normally distributed with mean value 65 and with a standard deviation of 5. If 3 students are selected at random from this set, what is the probability that exactly 2 of them will have marks over 70?

Solution:

Let X denotes marks obtained by a student in a certain subject. Then
$X \sim N(65,25)$ and

$$p(x) = P(X > 70) = P\left(Z > \frac{70-65}{5}\right) = P(Z > 1)$$

$$= 0.5 - P(0 < Z < 1)$$

$$= 0.5 - 0.3413 \qquad \text{(Use table)}$$

$$= 0.1587$$

Let Y denote the number of students who got more than 70 marks in a sample of 3 students.

Then $Y \sim B(3, p)$ and $P(Y = 2) = \binom{3}{2}(0.1587)^2(0.8413)$

P3:

The time taken by a person while speaking over a telephone is exponentially distributed with mean 4 *minutes.*

i) **Find the probability that he speaks for more than 6 *minutes* but less than 7 *minutes*.**

ii) **Out of 6 calls that he makes, what is the probability that exactly 2 calls take him more than 3 minutes each.**

iii) **How many calls out of 100 are expected to take more than 3 minutes each?**

Solution:

Let X be the time taken (in minutes) per call. We are given that X is exponentially distributed with mean 4 *minutes*.

$$\therefore f(x) = \begin{cases} \frac{1}{4} e^{-\frac{x}{4}} & , \quad x > 0 \\ 0 & , \quad otherwise \end{cases}$$

(i) $$P(6 < X < 7) = \frac{1}{4}\int_6^7 e^{-\frac{x}{4}} = \frac{1}{4}\left[\frac{e^{-\frac{x}{4}}}{-\frac{1}{4}}\right]_6^7 = e^{-\frac{6}{4}} - e^{-\frac{7}{4}} (= 0.04936)$$

(ii) $$P(X > 3) = \frac{1}{4}\int_3^\infty e^{-\frac{x}{4}} dx = \frac{1}{4}\left[\frac{e^{-\frac{x}{4}}}{-\frac{1}{4}}\right]_3^\infty = e^{-\frac{3}{4}} (= 0.4724)$$

Let Y denote the number of calls each with more than 3 minutes out of 6 calls. Then

$Y \sim B(6, P)$ where $P = 0.4724$

$$\therefore P(Y = 2) = \binom{6}{2} (0.4764)^2 (0.5276)^4 = 0.2594$$

(iii) Expected number of calls out of 100 that will be longer than 3 minutes each

$= 100 \times P(X > 3)$

$= 100 \times 0.4724 = 47.24 =$ 47(approximately)

P4:

The mileage (in thousands of miles) which car owners get with a certain kind of tyres is a random variable having probability density function

$$\boldsymbol{f(x) = \begin{cases} \frac{1}{10} e^{-\frac{x}{10}} & , \quad x > 0 \\ 0 & , \quad otherwise \end{cases}}$$

Find the probability that one of these tyres will last

- **i) at most 5,000 miles.**
- **ii) anywhere from 8,000 to 12,000 miles.**

Solution:

Let X represents the mileage in thousands of miles

i) $P(X \leq 5) = F(5) = 1 - e^{-\frac{5}{10}} = 1 - e^{-\frac{1}{2}} = 1 - 0.6065 = 0.3935$

ii) $P(8 < X < 12) = F(12) - F(8) = e^{-0.8} - e^{-1.2} = 0.148$

2.5. Continuous Probability Distributions

Exercise:

1) X is a normal variate with mean 30 and standard deviation 5. Find the probabilities that
   (i) $26 \leq X \leq 40$ (ii) $X \geq 45$ (iii) $|X - 30| > 5$

2) There are 600 engineering students in the B.Tech. classes of a university and the probability for any student to need a copy of a particular book from the university library on any day is 0.05. How many copies of the book should be kept in the university library so that the probability may be greater than 0.90 that none of the students needing a copy from the library has to come back disappointed? (use normal approximation to the binomial distribution)

3) In a distribution exactly normal, 10.03% of the items are under $25kg$ weight and 89.97% of the items are under $70kg$ weight. What are the mean and standard deviatioin of the distribution?

4) In an examination it is laid down that a student passes if he secures 30 percent or more marks. He is placed in the first, second or third division according as he secures 60% or more marks, between 45% and 60% marks and marks between 30% and 45% respectively. He gets distinction in case he secures 80% or more marks. It is noticed from the result that 10% of the students failed in the examination, where as 5% of them obtained distinction. Calculate the percentage of students placed in the examination. (Assume normal distribution to marks)

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 defectives and (ii) exactly 44 defectives? (use normal approximation to the binomial distribution)

6) If X is uniformaly distributed with mean 1 and variance $\frac{4}{3}$, find $P(X < 0)$.

7) Subway trains on a certain line run every half hour between mid-night and six in the morning. What is the probability that a man entering the station at a random time during this period will have to wait at least twenty minutes?

8) Suppose the life time of an electric component has exponential distribution with a mean life of $500\ hrs$.
   (i) Find the probability that it will give additional $600\ hrs$ life given that the component has been working for the last $300hrs$.
   (ii) Find the probability that it will work for more than $600hrs$.

9) The life(in hours) of electronic tubes manufactured by a certain process is known to have p.d.f.

$$f(x) = \begin{cases} \dfrac{1}{400} e^{-\frac{1}{400}(x-400)} & , \quad x \geq 400 \\ 0 & , \quad otherwise \end{cases}$$

   Find the probability of one such tube lasting for
   (i) at least $800hrs$.
   (ii) at most $1200hrs$.
   (iii) anywhere between 500 and $800hrs$.

Answers:

1) (i) 0.7653 (ii) 0.00135 (iii) 0.3174
2) 37
3) $\mu = 47.5$ and $\sigma = 17.578$
4) 34%
5) (i) 0.2376 (ii) 0.0584
6) $\frac{1}{4}$
7) $\frac{1}{3}$
8) (i) $e^{-\frac{6}{5}}$ (ii) $e^{-\frac{6}{5}}$
9) (i) e^{-1} (ii) e^{-2} (iii) $e^{-\frac{1}{4}} - e^{-1}$

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 $(X, Y)\colon S \longrightarrow R \times R$, $i.e., (X,Y)(\omega) = \big(X(\omega), Y(\omega)\big)$ $\forall\, \omega \in S$, 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 $\sum\limits_{i=1}^{m} \sum\limits_{j=1}^{n} p\left(x_i, y_j\right) = 1$. 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

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

$$p_{2|1}(\,y_j |\, x_i) = \frac{p(x_i, y_j)}{p_1(x_i)} \quad j = 1,2,3, \dots, n$$

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:

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

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

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

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

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

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

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

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

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

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

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

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

$$f_1(x)=\int_0^\infty f(x,y)dy=\int_0^\infty e^{-(x+y)}dy=e^{-x}\int_0^\infty e^{-y}dy=e^{-x}.1=e^{-x}$$

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

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

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

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

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

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

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

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

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

$\Longrightarrow f_{1|2}(x|y) = x(y+1)^2 e^{-x.(y+1)}$ 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 $f_1(x)f_2(y) = \frac{x^3}{4}.2y = \frac{x^3y}{2} = f(x,y), \forall(x,y)$

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

P1:

The c.d.f. of bivariate discrete random variable (x, y) is given by

$$F(x,y)=\begin{cases}\frac{1}{8} & for \;\; x=1, y=1\\ \frac{5}{8} & for \;\; x=1, y=2\\ \frac{1}{4} & for \;\; x=2, y=1\\ 1 & for \;\; x=2, y=2\end{cases}$$

Find a) j.p.m.f. of (x, y) b) m.p.m.f. of x c) m.p.m.f. of y

Solution:

The j.p.m.f. is obtained from the relationship $F(x,y)=\sum_{t\le x}\sum_{s\le y} p(t,s)$. Thus

$$F(1,1) = \frac{1}{8} = p(1,1)$$

$$F(1,2) = p(1,1) + p(1,2) = \frac{5}{8} \Rightarrow p(1,2) = \frac{5}{8} - \frac{1}{8} = \frac{1}{2}$$

$$F(2,1) = p(1,1) + p(2,1) = \frac{1}{4} \Rightarrow p(2,1) = \frac{1}{4} - \frac{1}{8} = \frac{1}{8}$$

$$F(2,2) = p(1,1) + p(1,2) + p(2,1) + p(2,2) = 1 \Rightarrow p(2,2) = \frac{1}{4}$$

Thus, the j.p.m.f is given by

$$p(x,y)=\begin{cases}\frac{1}{8} & for \; x=1, y=1\\ \frac{1}{2} & for \; x=1, y=2\\ \frac{1}{8} & for \; x=2, y=1\\ \frac{1}{4} & for \; x=2, y=2\end{cases}$$

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

$$p_1(x) = \begin{cases} p(1,1) + p(1,2) = \dfrac{5}{8} & for \;\; x = 1 \\ p(2,1) + p(2,2) = \dfrac{3}{8} & for \;\; x = 2 \end{cases}$$

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

$$p_2(y) = \begin{cases} p(1,1) + p(2,1) = \dfrac{1}{4} & for \;\; y = 1 \\ p(1,2) + p(2,2) = \dfrac{3}{4} & for \;\; y = 2 \end{cases}$$

P2:

The j.p.m.f. of (X, Y) is given in the following table:

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

Find a) The m.p.m.fs of x and y

b) Conditional p.m.f. of x given $y = 1$

c) Conditional p.m.f. of y given $x = 2$

Solution:

a) The marginal p.m.f of x is given by

$$p_1(x) = \sum_y p(x, y)$$

$$\therefore p_1(1) = \sum_{y=1}^{4} p(1, y) = p(1,1) + p(1,2) + p(1,3) + p(1,4)$$

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

Similarly,

$$p_1(2) = \sum_{y=1}^{4} p(2, y) = \frac{9}{36}$$

$$p_1(3) = \sum_{y=1}^{4} p(3, y) = \frac{8}{36} \quad and$$

$$p_1(4) = \sum_{y=1}^{4} p(4, y) = \frac{9}{36}$$

Thus, the m.p.m.f. of x is given in the following table:

x	1	2	3	4
$p_1(x)$	$\frac{10}{36}$	$\frac{9}{36}$	$\frac{8}{36}$	$\frac{9}{36}$

Similarly, we can obtain the m.p.d.f. of y as given in the following table:

y	1	2	3	4
$p_2(y)$	$\frac{11}{36}$	$\frac{9}{36}$	$\frac{7}{36}$	$\frac{9}{36}$

b) The conditional p.m.f. of x given $y = 1$ is given by

$p_{1|2}(x|1) = \frac{p(x,1)}{p_2(1)}$ for $x = 1,2,3,4.$

$p_{1|2}(1|1) = \frac{p(1,1)}{p_2(1)} = \frac{\frac{4}{36}}{\frac{11}{36}} = \frac{4}{11}.$

Similarly, we can find

$p_{1|2}(2|1) = \frac{1}{11}$, $p_{1|2}(3|1) = \frac{5}{11}$ and $p_{1|3}(4|1) = \frac{1}{11}$

Hence, the conditional p.m.f. of x given $y = 1$ is given in the following table:

x	1	2	3	4
$p_{1|2}(x|1)$	$\frac{4}{11}$	$\frac{1}{11}$	$\frac{5}{11}$	$\frac{1}{11}$

Similarly, we can obtain the conditional p.m.f. of y given $x = 2$ as given in the following table:

y	1	2	3	4
$p_{2|1}(y|2)$	$\frac{1}{9}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{2}{9}$

P3:

If X and Y are two random variables having j.p.d.f.

$$f(x,y)=\begin{cases}\frac{1}{8}(6-x-y), & 0<x<2, \quad 2<y<4\\ 0 \quad , & otherwise\end{cases}$$

Find (i) $P(X<1\cap Y<3)$ (ii) $P(X+Y<3)$ and (iii) $P(X<1|Y<3)$.

Solution:

$$(i)\ P(X<1\cap Y<3)=\int_0^1\int_0^3 f(x,y)\,dxdy=\frac{1}{8}\int_0^1\int_2^3(6-x-y)\,dxdy=\frac{3}{8}$$

$$(ii)\ P(X+Y<3)=\frac{1}{8}\int_0^1\int_2^{3-x}(6-x-y)\,dxdy=\frac{5}{24}$$

$$(iii)\ P(X<1|Y<3)=\frac{P(x<1\cap y<3)}{P(y<3)}$$

$$But\ P(Y<3)=\frac{1}{8}\int_0^2\int_2^3(6-x-y)\,dxdy=\frac{5}{8}$$

$$Hence\ P(X<1|Y<3)=\frac{\frac{3}{8}}{\frac{5}{8}}=\frac{3}{5}$$

P4:

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

$$f(x,y)=\begin{cases}2, & 0<x<1, 0<y<x\\ 0, & otherwise\end{cases}$$

a) **Find marginal and conditional p.d.fs of x and y.**
b) **Check for independence of x and y.**

Solution:

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

$$f_1(x)=\int_0^x f(x,y)dy=\int_0^x 2dy=2x$$

$$\Rightarrow f_1(x)=2x, 0<x<1$$

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

$$f_2(y)=\int_y^1 f(x,y)dx=\int_y^1 2dx=2(1-y)$$

$$\Rightarrow f_2(y)=2(1-y), 0<y<1$$

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

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

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

$$f_{2|1}(y|x)=\frac{f(x,y)}{f_1(x)}=\frac{2}{2x}=\frac{1}{x}, 0<x<1$$

b) Since $f_1(x)\cdot f_2(y)=2x\cdot 2(1-y)=4x(1-y)\neq f(x,y)$, X and Y are not 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

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

provided these values exist.

Mean and moments:

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

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

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

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

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 $g(X) = \left(X - E(X)\right)^r = (X-\mu)^r$. 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 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 $E(X)$ and $E(X^2)$.

Solution:

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

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

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 $X = 2,3,4, \dots ,12$ 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
$p(x)$	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$

$$\mathrm{E(X)} = \sum_x x\, p(x) = 0 \times \frac{1}{16} + 1 \times \frac{4}{16} + 2 \times \frac{6}{16} + 3 \times \frac{4}{16} + 4 \times \frac{1}{6}$$

$$= \frac{1}{16}(0 + 4 + 12 + 12 + 4) = \frac{32}{16} = 2$$

$$\Rightarrow \mathrm{E(X)} = 2$$

$$V(X) = E(X-2)^2 = \sum (x-2)^2\, p(x)$$

$$= (0-2)^2 \times \frac{1}{16} + (1-2)^2 \times \frac{4}{16} + (2-2)^2 \times \frac{6}{16} + (3-2)^2 \times \frac{4}{16} + (4-2)^2 \times \frac{1}{16}$$

$$= 4 \times \frac{1}{16} + 1 \times \frac{4}{16} + 0 \times \frac{6}{16} + 1 \times \frac{4}{16} + 4 \times \frac{1}{16} = \frac{1}{16}(4+4+4+4) = \frac{16}{16} = 1$$

$\Longrightarrow \mathrm{V(X)} = 1$

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

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

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 $X_1, X_2, X_3 \dots X_n$ are random variables, then $E(X_1 + X_2 + X_3 + \cdots + X_n) = E(X_1) + E(X_2) + E(X_3) \dots + E(X_n)$ 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,

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

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 $X_1, X_2, X_3 \dots X_n$ are independent random variables, then $E(X_1 X_2 X_3 \dots . X_n) = E(X_1)E(X_2)E(X_3) \dots E(X_n)$.

Theorem 4: Mathematical expectation of a linear combination of random variables.

Let $X_1, X_2, X_3 \ldots X_n$ be any n random variables and $a_1, a_2, a_3, \ldots . a_n$ be any n constants. Then

$$E(a_1X_1 + a_2X_2 + a_3X_3 + \cdots + a_nX_n) = a_1E(X_1) + a_2E(X_2) + a_3E(X_3) + \cdots + a_nE(X_n)$$

provided all the expectations exist.

The proof follows using Theorem 1 and generalization of Theorem 2.

Theorem 5: $V(X) = E(X^2) - \left(E(X)\right)^2$

Proof: $V(X) = E[X - E(X)]^2$

$$= E\left[X^2 - 2XE(X) + \left(E(X)\right)^2\right]$$

$$= E(X^2) - 2E\big(X\,E(X)\big) + E\big(E(X)\big)^2$$

$= E(X^2) - 2E(X)E(X) + \left(E(X)\right)^2 \quad \because E(X)$ 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 $E\big(X - E(X)\big)^2$.

Theorem 6: If X is a random variable, and a and b are constants, then $V(ax + b) = a^2\, V(X)$.

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 $\boldsymbol{a = 0}$, then $\boldsymbol{V(b) = 0}$ $\boldsymbol{i.e.}$, variance of a constant is zero.

Corollary 2: If $\boldsymbol{b = 0}$, then $\boldsymbol{V(aX) = a^2V(X)}$

Covariance: If X and Y are two random variables, then the **covariance** between them is defined by

$$\text{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))]$$

$$= E[XY - XE(Y) - YE(X) + E(X)E(Y)]$$

$$= E(XY) - E\left((XE(Y)) - E(YE(X)\right) + E(E(X)E(Y))$$

$$= E(XY) - E(X)E(Y) - E(Y)E(X) + E(X)E(Y)$$

$$= E(XY) - \ E(X)E(Y)$$

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 $\boldsymbol{X_1, X_2, X_3, \dots, X_n}$ be any $\boldsymbol{n}$ random variables and $\boldsymbol{a_1, a_2, a_3, \dots, a_n}$ are $\boldsymbol{n}$ constants, then

$$\boldsymbol{V\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 V(X_i) + 2\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j Cov(X_i, X_j)}$$

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 $U - E(U) = \sum_{i=1}^{n} a_i \left(X - E\left(X_i\right)\right)$

$$\Rightarrow \left(U - E(U)\right)^2 = \sum_{i=1}^{n} a_i^2\left(X_i - E\left(X_i\right)\right)^2 + 2\sum_{\substack{i=1 \\ i<j}}^{n}\sum_{j=1}^{n} a_i a_j \left(X_i - E\left(X_i\right)\right)\left(X_j - E\left(X_j\right)\right)$$

$$\Rightarrow E\left(U - E(U)\right)^2 = \sum_{i=1}^{n} a_i^2 E\left(X_i - E\left(X_i\right)\right)^2 + 2\sum_{\substack{i=1 \\ i<j}}^{n}\sum_{j=1}^{n} a_i a_j E\left[\left(X_i - E\left(X_i\right)\right)\left(X_j - E\left(X_j\right)\right)\right]$$

$$\Rightarrow V\left(\sum_{i=1}^{n} a_i X_i\right) = V(U) = \sum_{i=1}^{n} a_i^2 V\left(X_i\right) + 2\sum_{\substack{i=1 \\ i<j}}^{n}\sum_{j=1}^{n} a_i a_j \, cov\left(X_i, X_j\right)$$

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** $V(X_1 + X_2) = V(X_1) + V(X_2) + 2Cov(X_1, X_2)$
3. **If $a_1 = 1, a_2 = -1$ and $a_3 = \cdots . = a_n = 0$, then** $V(X_1 - X_2) = V(X_1) + V(X_2) - 2Cov(X_1, X_2)$
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!)

$$\therefore Cov(X,Y) = E(XY) - E(X)E(Y) = \frac{1}{6} - \frac{5}{12}.\frac{5}{12} = -\frac{1}{144}$$

P1:

Let X be a d.r.v. with p.m.f given by

$$p(x)=\begin{cases}\frac{1}{3}, & x=0\\ \frac{2}{3}, & x=2\end{cases}$$

Find the mean and variance of X.

Solution:

Here $p(0)=\frac{1}{3}$ and $p(2)=\frac{2}{3}$

Then

$$E(X)=\sum_x x\,p(x)=0\times\frac{1}{3}+2\times\frac{2}{3}=\frac{4}{3}$$

and

$$E(X^2)=\sum_x x^2\,p(x)=0^2\times\frac{1}{3}+2^2\times\frac{2}{3}=\frac{8}{3}$$

$$\therefore V(X)=E(X^2)-\left(E(X)\right)^2=\frac{8}{3}-\frac{16}{9}=\frac{8}{9}$$

Thus, mean is equal to $\frac{4}{3}$ and variance is equal to $\frac{8}{9}$.

P2:

Find the mean and variance of the random variable, whose p.d.f. is given by

$$f(x)=\begin{cases}\lambda e^{-\lambda x} & ,x>0\\ 0 & ,otherwise\end{cases}$$

Solution:

$$E(X)=\int_{-\infty}^{\infty} x\,f(x)dx=\int_{0}^{\infty} x\,f(x)dx=\lambda\int_{0}^{\infty} x\,e^{-\lambda x}dx=\lambda\left[\left(-\frac{xe^{-\lambda x}}{\lambda}\right)_0^{\infty}+\frac{1}{\lambda}\int_0^{\infty} e^{-\lambda x}dx\right]$$

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

and $E\left(X^2\right)=\int_{-\infty}^{\infty} x^2 f(x)dx=\int_{0}^{\infty} x^2 f(x)dx=\lambda\int_{0}^{\infty} x^2 e^{-\lambda x}dx$

$$=\lambda\left[\left(\frac{x^2.e^{-\lambda x}}{\lambda}\right)_0^{\infty}+2\int_0^{\infty}\frac{x.e^{-\lambda x}}{\lambda}dx\right]=\lambda\left[0+\frac{2}{\lambda}\int_0^{\infty} x.e^{-\lambda x}dx\right]=\lambda\left[\frac{2}{\lambda}.\frac{1}{\lambda^2}\right]=\frac{2}{\lambda^2}$$

Thus $V(X)=E(X^2)-\left(E(X)\right)^2=\frac{2}{\lambda^2}-\frac{1}{\lambda^2}=\frac{1}{\lambda^2}$

P3:

A student doing a summer internship in a company was asked to model the life term of certain equipment that the company makes. After a series of tests, the student proposed that the life time of the equipment can be modeled by a random variable X that has p.d.f.

$$f(x) = \begin{cases} \dfrac{x e^{-\frac{x}{10}}}{100}, & x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

a. **Show that $f(x)$ is a valid p.d.f.**
b. **What is the probability that the lifetime of the equipment exceeds 20?**
c. **Find the mean?**

Solution:

a. Consider $\int\limits_0^\infty f(x)dx = \dfrac{1}{100}\int\limits_0^\infty x e^{-\frac{x}{10}} dx$

$$= \frac{1}{100}\left\{\left[-10xe^{-\frac{x}{10}}\right]_0^\infty + 10\int\limits_0^\infty e^{-\frac{x}{10}} dx\right\}$$

$$= \frac{1}{100}\left\{0 - 100\left[e^{-\frac{x}{10}}\right]_0^\infty\right\} = \frac{1}{100} \times 100 = 1$$

Thus, $f(x)$ is a valid p.d.f.

b. $P(X > 20) = \int\limits_{20}^\infty f(x)dx = \dfrac{1}{100}\left\{\left[-10xe^{-\frac{x}{10}}\right]_{20}^\infty + 10\int\limits_{20}^\infty e^{-\frac{x}{10}} dx\right\}$

$$=\frac{1}{100}\left\{200e^{-2}-\left[100e^{-\frac{x}{10}}\right]_{20}^{\infty}\right\}=2e^{-2}+e^{-2}=3e^{-2}=\frac{3}{e^2}=0.406$$

c. Mean= $E(X)=\int_0^\infty x\,f(x)dx=\frac{1}{100}\int_0^\infty x^2e^{-\frac{x}{10}}dx$

$$=\frac{1}{100}\left[10x^2e^{-\frac{x}{10}}\right]_0^\infty+\frac{20}{100}\int_0^\infty xe^{-\frac{x}{10}}dx`$$

$= 0 + 20 = 20$

P4:

A box contains a white and b black balls. c balls are drawn at random. Find the expected value of the number of white balls drawn.

Solution:

Let a variable X_i associated with i^{th} drawn, be defined as follows:

$$X_i = \begin{cases} 1, \text{if the } i^{th} \text{ ball drawn is white} \\ 0, \text{if the } i^{th} \text{ ball drawn is black} \end{cases}$$

Then the number of S of white balls among c balls drawn is given by

$$S = X_1 + X_2 + \cdots + X_c = \sum_{i=1}^{c} X_i$$

$$\Rightarrow E(S) = \sum_{i=1}^{c} E(X_i)$$

But $P(X_i = 1) = P(\text{ of drawing a white ball }) = \frac{a}{a+b}$

and $P(X_i = 0) = P(\text{ of drawing a black ball }) = \frac{b}{a+b}$

$\therefore E(X_i) = 1.P(X_i = 1) + 0.P(X_i = 0) = P(X_i = 1) = \frac{a}{a+b}$ and hence

$$E(S) = \sum_{i=1}^{c} E(X_i) = \sum_{i=1}^{c} \frac{a}{a+b} = \frac{ac}{a+b}.$$

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

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$$

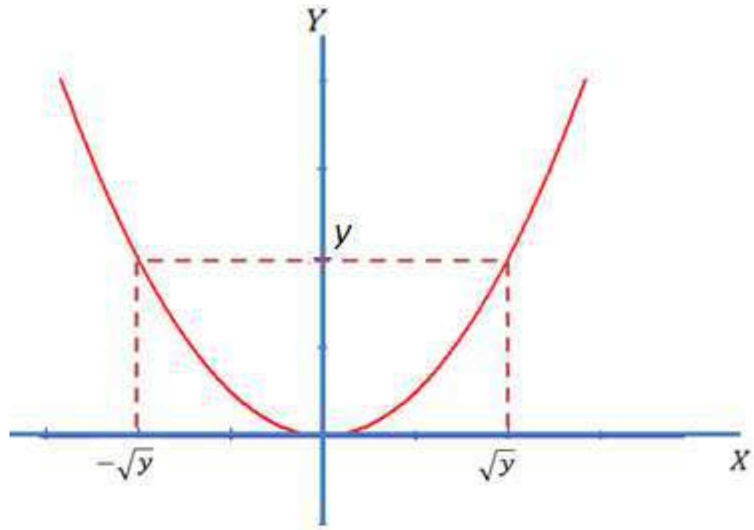
Example 1: Find the p.d.f of Y in terms of the p.d.f of X if $Y = 2X + 7$.

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

$$F_Y(y) = P[Y \le y] = P[X^2 \le y]$$

$$= P\left[|X| \le \sqrt{y}\right], \quad y > 0$$

$$= P\left[-\sqrt{y} \le X \le \sqrt{y}\right]$$

$$= F_X\left(\sqrt{y}\right) - F_X\left(-\sqrt{y}\right)$$

The p.d.f of Y is given by

$$f_Y(y) = \frac{dF_Y(y)}{dy} = \frac{d}{dy}\left[F_X\left(\sqrt{y}\right) - F_X\left(-\sqrt{y}\right)\right]$$

Let $u = \sqrt{y} = y^{\frac{1}{2}}$. Thus, $\frac{du}{dy} = \frac{1}{2}y^{-\frac{1}{2}}$ and

$$f_Y(y) = \frac{d}{dy}\left[F_X\left(\sqrt{y}\right) - F_X\left(-\sqrt{y}\right)\right]$$

$$= \frac{d}{du}\left(F_X(u)\right)\frac{du}{dy} + \frac{d}{du}\left(F_X(-u)\right)\frac{du}{dy}$$

$$= \frac{1}{2} y^{-\frac{1}{2}} \left[\frac{d}{du}\left(F_X(u)\right) + \frac{d}{du}\left(F_X(-u)\right)\right]$$

$$= \frac{1}{2} y^{-\frac{1}{2}} \left[f_X\left(\sqrt{y}\right) + f_X\left(-\sqrt{y}\right)\right]$$

$$= \frac{f_X(\sqrt{y}) + f_X(-\sqrt{y})}{2\sqrt{y}} \quad , y > 0$$

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

$$f_Y(y) = \frac{f_X(\sqrt{y}) + f_X(-\sqrt{y})}{2\sqrt{y}} = \frac{2 f_X(\sqrt{y})}{2\sqrt{y}} = \frac{f_X(\sqrt{y})}{\sqrt{y}}$$

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

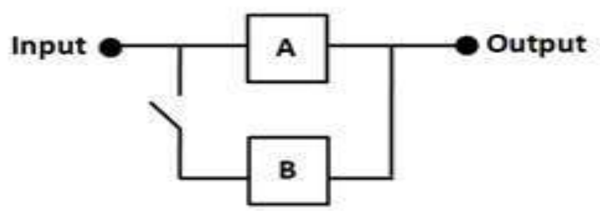
$$f_Y(y) = \frac{dF_Y(y)}{dy} = 2\frac{dF_X(u)}{du}\frac{1}{2\sqrt{y}}$$

$$= \frac{1}{\sqrt{y}} f_X\left(\sqrt{y}\right)$$

$$= \frac{1}{\sqrt{2\pi y}} e^{-\frac{y}{2}}, \; y > 0$$

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,

$$F_S(s) = \int_{-\infty}^{\infty}\int_{-\infty}^{s-y} f_{XY}(x,y)\,dxdy = \int_{-\infty}^{\infty}\int_{-\infty}^{s-y} f_X(x)\,f_Y(y)\,dxdy$$

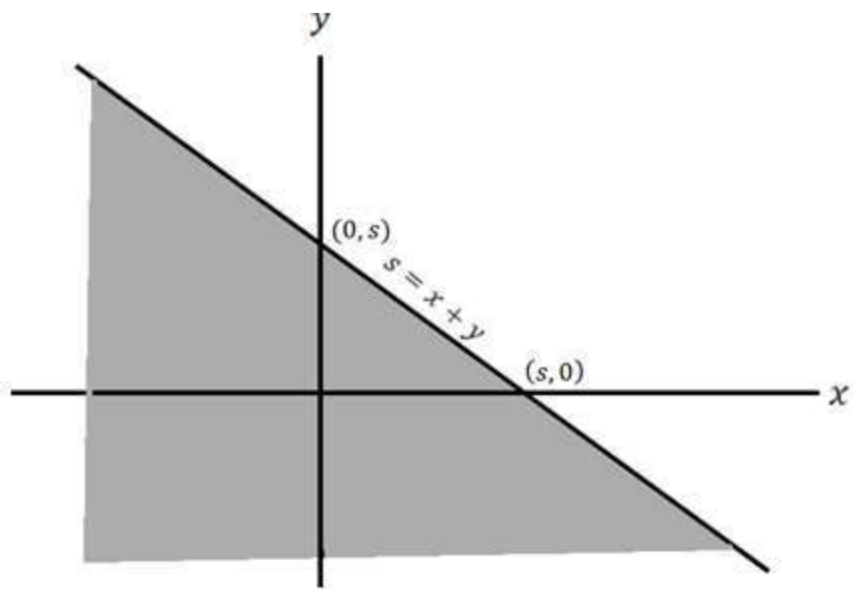
$$= \int_{-\infty}^{\infty}\left\{\int_{-\infty}^{s-y} f_X(x)\,dx\right\} f_Y(y)\,dy$$

$$= \int_{-\infty}^{\infty} F_X(s-y)\,f_Y(y)\,dy$$

The p.d.f. of S is obtained by differentiating the c.d.f. , as follows:

$$
\begin{aligned}
f_S(s) &= \frac{d}{ds} F_S(s) = \frac{d}{ds} \int_{-\infty}^{\infty} F_X(s-y)\, f_Y(y)\, dy \\
&= \int_{-\infty}^{\infty} \frac{d}{ds} F_X(s-y)\, f_Y(y)\, dy \\
&= \int_{-\infty}^{\infty} f_X(s-y)\, f_Y(y)\, dy
\end{aligned}
$$

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



$D = \{(x, y) | x + y \leq s\}$

$fig. 2$

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.

$$\boldsymbol{f_X(u) = f_Y(u)} = \begin{cases} \mathbf{\frac{1}{4}} & \mathbf{0} < u < 4 \\ \mathbf{0} & \mathbf{otherwise} \end{cases}$$

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),

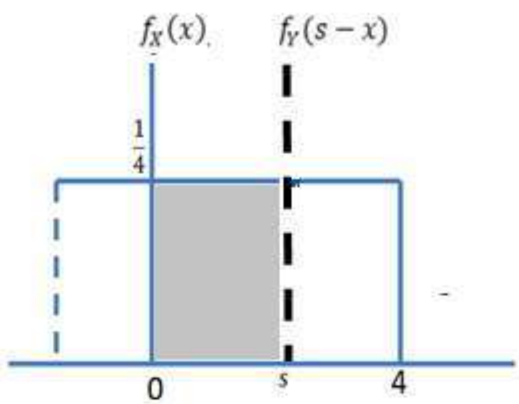
$$f_S(s) = \int_0^s \frac{1}{16} dy = \frac{s}{16}$$

For $4 < s < 8$ (see $fig.3\ (b)$), we obtain

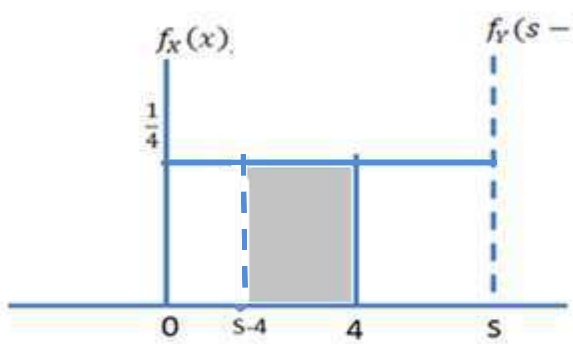
$$f_S(s) = \int_{s-4}^4 dy = \frac{8-s}{16}$$

Thus ,

$$f_S(s) = \begin{cases} \frac{s}{16} & ,\ 0 \le s \le 4 \\ \frac{8-s}{16} & ,\ 4 < s < 8 \\ 0 & ,\ \text{otherwise} \end{cases}$$



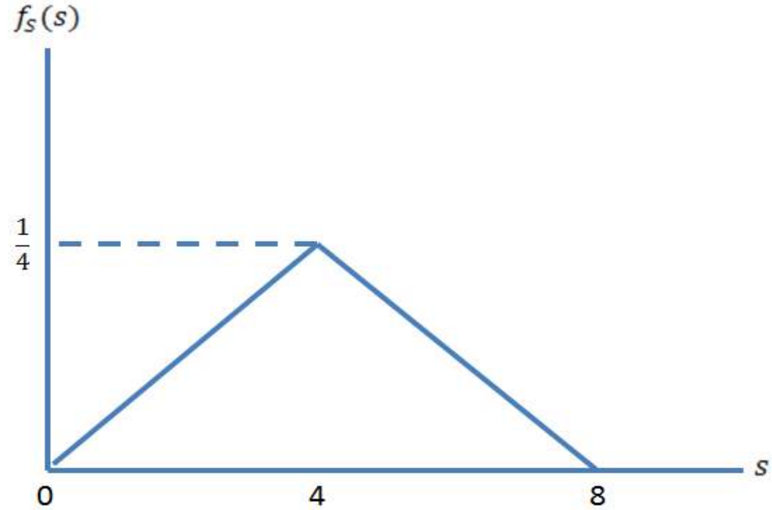
$fig\ 3(a)$



$fig\ 3(b)$

$Fig.3$: Convolution of p.d.fs (a) $0 \le s \le 4$ and (b) $4 \le s \le 8$

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:

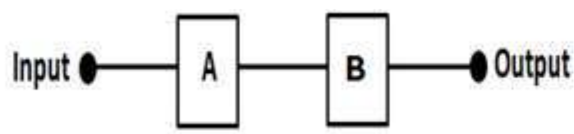
$$F_U(u) = P[U \leq u] = P[\min(X,Y) \leq u] = P[(X \leq u, X \leq Y) \cup (Y \leq u, X > Y)]$$

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

$$f_U(u) = f_X(u)\{1 - F_Y(u)\} + f_Y(u)\{1 - F_X(u)\}$$

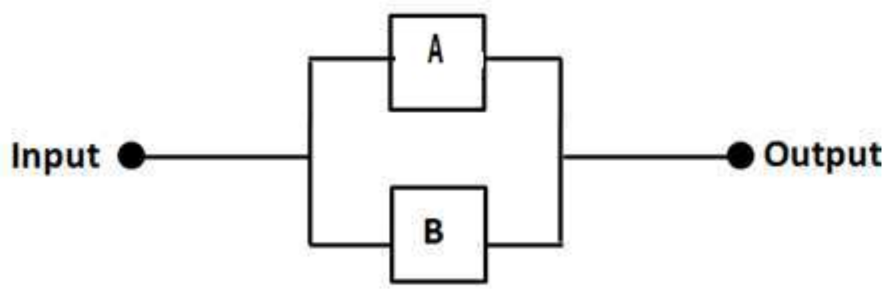
$$= \lambda e^{-\lambda u} e^{-\mu u} + \mu e^{-\mu u} e^{-\lambda u}$$

$$= (\lambda + \mu)e^{-(\lambda+\mu)u} \ , \ u \geq 0$$

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 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ 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

$$f_{UW}(u, w) = \frac{f_{XY}(x, y)}{|J(x, y)|} = \frac{1}{|w|} f_{XY}\left(w, \frac{u}{w}\right)$$

$$f_U(u) = \int_{-\infty}^{\infty} f_{UW}(u, w) dw = \int_{-\infty}^{\infty} \frac{1}{|w|} f_{XY}\left(w, \frac{u}{w}\right) dw$$

P1:

Obtain the p.d.f. of $Z = X + Y$, where X and Y are two independent random variables with the following p.d.fs:

$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & ; \quad a < x < b \\ 0 & ; \quad otherwise \end{cases}$$

$$f_Y(y) = \begin{cases} \dfrac{1}{d-c} & ; \quad c < y < d\,; \;\; d - c < b - a \\ 0 & ; \quad otherwise \end{cases}$$

Solution:

The two p.d.fs are shown in the *fig.* 1. To evaluate the limits of integration of the p.d.f. of Z, we consider the following regions represented by the diagram shown in *fig.* 2.

When $z < a + c$, $f_Z(z) = 0$. When $a + c \le z \le a + d$ (see *figure 2(i)*, we obtain)

$$f_Z(z) = \frac{1}{(b-a)(d-c)} \int_a^{z-c} dy = \frac{z-c-a}{(b-a)(d-c)}$$

When $a + d \le z \le \mathrm{b} + \mathrm{c}$ (see *fig.* $2(ii)$), we obtain)

$$f_Z(z) = \frac{1}{(b-a)(d-c)} \int_{z-d}^{z-c} dy = \frac{1}{b-a}$$

When $b + c \le z \le b + d$ (see *fig.* $2(ii)$), we obtain

$$f_Z(z) = \frac{1}{(b-a)(d-c)} \int_{z-d}^{b} dy = \frac{b+d-z}{(b-a)(d-c)}$$

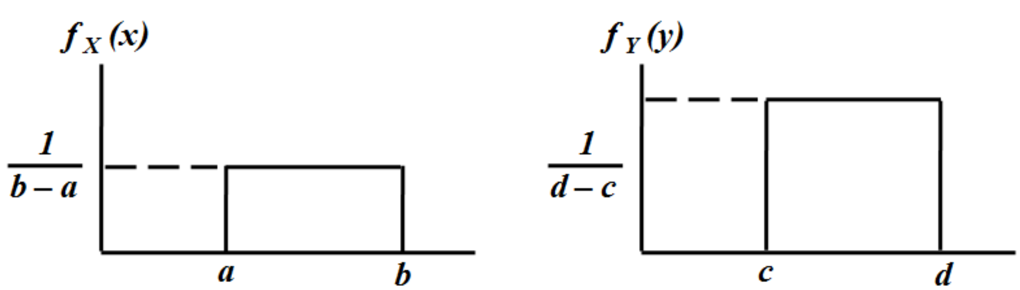


Figure 1: p.d.fs of X and Y

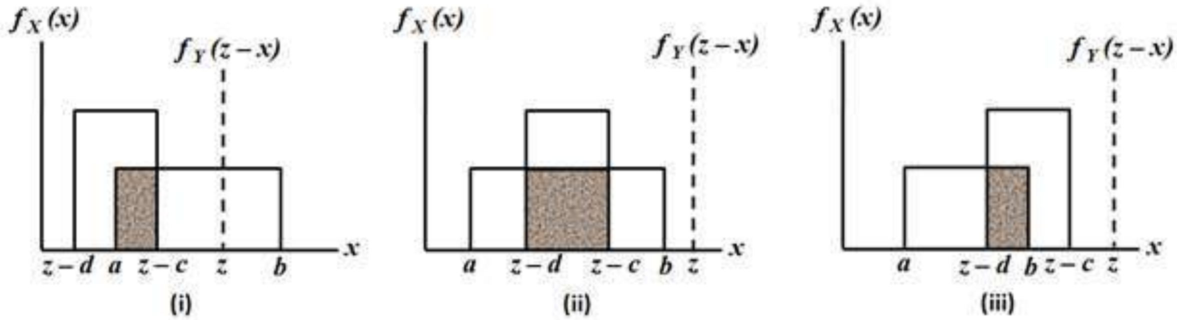
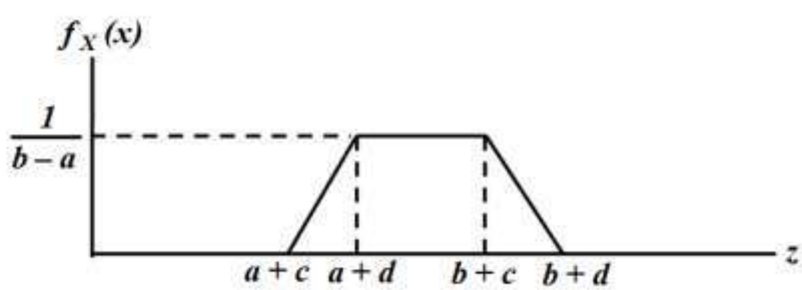


Figure 2: Convolution of the p.d.fs for different values z

Finally, when $z > b + d$, $f_Z(z) = 0$, thus, the p.d.f of Z is given by

$$f_Z(z) = \begin{cases} 0 & , & z < a + c \\ \dfrac{z - a - c}{(b - a)(d - c)} & , & a + c \leq z \leq a + d \\ \dfrac{1}{b - a} & , & a + d \leq z \leq b + c \\ \dfrac{b + d - z}{(b - a)(d - c)} & , & b + c \leq z \leq b + d \\ 0 & , & z > b + d \end{cases}$$

The p.d.f is graphically illustrated in the following figure which is a **trapezoid**. Note that when $b - a = d - c$, the p.d.f reduces to an **isosceles triangle** centered at $z = \frac{a+c+b+d}{2}$. In the special case when $a = c$ and $b = d$, the isosceles triangle is centered at $z = a + b$.



P2:

Find the p.d.f of U, which is the sum of X and Y that are independent random variables with the following p.d.fs:

$$\boldsymbol{f_X(x) = \lambda e^{-\lambda x} \quad , \quad x \geq 0}$$
$$\boldsymbol{f_Y(y) = \lambda^2 y e^{-\lambda y} \quad , \quad y \geq 0}$$

Solution:

Since X and Y are independent random variables and the p.d.f of U is given by

$$f_U(u) = \int_0^u f_X(x) f_Y(u-x)\, dx$$
$$= \int_0^u \lambda e^{-\lambda x}\, \lambda^2 (u-x) e^{-\lambda(u-x)} dx$$
$$= \lambda^3 e^{-\lambda u} \int_0^u (u-x)\, dx$$
$$= \lambda^3 e^{-\lambda u} \left[ux - \frac{x^2}{2}\right]_0^u = \lambda^3 e^{-\lambda u} \left[u^2 - \frac{u^2}{2}\right]$$
$$= \frac{\lambda^3 u^2 e^{-\lambda u}}{2}$$
$$= \frac{\lambda^3 u^2 e^{-\lambda u}}{2!} \quad u \geq 0$$

This is known as **Erlang – 3 distribution**.

2.6. Functions of Random Variables:

Exercise

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

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

Find the probability distribution of $Y = 4X + 3$.

2) A r.v. X is exponentially distributed with parameter1. Find the p.d.f. of $y = \sqrt{x}$

3) Let X be a c.r.v with p.d.f.

$$f(x) = \begin{cases} \frac{x}{12} & , \quad 1 < x < 5 \\ 0 & , \quad otherwise \end{cases}$$

find the p.d.f. of $y = 2X + 4$

4) If the c.r.v X has the p.d.f.

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

find the p.d.f. of $y = X^2$

5) Find the $p.d.f$ of $W = X + Y$ where X and Y are independent r.v s with the following p.d.f s:

$$f_X(x) = \lambda e^{-\lambda x}, x \geq 0$$

and $\quad f_Y(y) = \mu e^{-\mu y}, \ y \geq 0,$ where $\lambda \neq \mu$.

6) The j.p.d.f of two random variables X and Y is given by $f_{XY}(xy)$. Find the p.d.f of $V = \frac{X}{Y}$

ANSWERS

1.

Y	-5	-1	7	11
$p(y)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

2. $f(y) = \begin{cases} 2y\, e^{-y^2}, & y \geq 0 \\ 0 \quad , & otherwise \end{cases}$

3. $f(y) = \begin{cases} \frac{y-4}{48}\ , & 6 < y < 14 \\ 0\ , & otherwise \end{cases}$

4. $f(y) = \begin{cases} \frac{2}{9\sqrt{y}} & ,0 < y < 1 \\ \frac{1}{9}\left(\frac{\sqrt{y}+1}{\sqrt{y}}\right) & ,1 < y < 4 \\ 0 & ,\ otherwise \end{cases}$

5. $f_W(w) = \begin{cases} \frac{\lambda\mu}{\lambda-\mu}\left(e^{-\mu w} - e^{-\lambda w}\right), & w \geq 0 \\ 0, & otherwise \end{cases}$

6. $f_V(v) = \int_{-\infty}^{\infty} |w|\, f_{XY}(vw, w)dw$

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 $\boldsymbol{\rho(X,Y)}$ or $\boldsymbol{\rho_{XY}}$, is a numerical measure of linear relationship between them and is defined by

$$\boldsymbol{\rho(X,Y) = \frac{\sigma_{XY}}{\sigma_X.\sigma_Y} = \frac{cov\ (X,Y)}{\sqrt{V(X)}\,\sqrt{V(Y)}}}$$

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 $X \sim N(0,1)$ and $Y = X^2$. Then $E(X) = E(X^3) = 0$.

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, $f_1(x) = \begin{cases} \frac{3}{2} - x & , \quad if \;\; 0 < x < 1 \\ 0 & , \quad otherwise \end{cases}$

Consider.

$$E(X)=\int_0^1 xf_1(x)dx=\int_0^1 x\left(\frac{3}{2}-x\right)dx=\int_0^1\left(\frac{3}{2}x-x^2\right)dx=\frac{5}{12}$$

$$E(X^2)=\int_0^1 xf_1(x)=\int_0^1 x^2\left(\frac{3}{2}-x\right)dx=\int_0^1\left(\frac{3}{2}x^2—x^3\right)dx=\frac{1}{4}$$

Further,

$$E(XY)=\int_0^1\int_0^1 xy\,f(x,y)dx\,dy\ =\int_0^1\int_0^1 xy\,(2-x-y)dx\,dy$$

$$=\int_0^1 y\left(\int_0^1(2x-x^2-xy)dx\ \right)dy=\int_0^1 y\left[2.\frac{x^2}{2}-\frac{x^3}{3}-\frac{yx^2}{2}\right]_0^1 dy$$

$$=\int_0^1 y\left(1-\frac{1}{3}-\frac{y}{2}\right)dy=\int_0^1 y\left(\frac{2}{3}-\frac{y}{2}\right)dy$$

$$=\int_0^1\left(\frac{2}{3}y-\frac{y^2}{2}\right)dy=\left[\frac{y^3}{3}-\frac{y^3}{6}\right]_0^1=\frac{1}{3}-\frac{1}{6}=\frac{1}{6}$$

$\therefore E(XY)=\frac{1}{6}$

Thus, $V(X)=E(X^2)-\big(E(X)\big)^2=\frac{1}{4}-\left(\frac{5}{12}\right)^2=\frac{1}{4}-\frac{25}{144}=\frac{36-25}{144}=\frac{11}{144}$

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 $(\boldsymbol{X}, \boldsymbol{Y}) \sim \boldsymbol{BN}\ (\mathbf{5}, \mathbf{10}, \mathbf{1}, \mathbf{25}, \boldsymbol{\rho})$ where $\boldsymbol{\rho} > 0$, find $\boldsymbol{\rho}$ when $\boldsymbol{P}(\mathbf{4} < Y < 16|X = 5) = \mathbf{0.954}$

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

$$\boldsymbol{f(x,y) = \frac{1}{2\pi\sqrt{3}} exp\left[-\frac{\{(2x-y)^2+2xy\}}{6}\right]}, -\infty < x, y < \infty$$

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

$$f(x,y) = \frac{1}{2\pi\sqrt{3}} exp\left[-\frac{\{(2x-y)^2+2xy\}}{6}\right], -\infty < x, y < \infty$$

$$i.e, \quad f(x,y) = \frac{1}{2\pi\sqrt{3}} exp\left[-\frac{\{4x^2+y^2-2xy\}}{6}\right] ----(\mathbf{2})$$

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

$$\boldsymbol{f(x,y) = c\, exp\left[-\frac{\{16(x-2)^2 - 12(x-2)(y+3) + 9(y+3)^2\}}{216}\right]}$$

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, $c=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}=\frac{1}{2\pi\times 3\times 4\sqrt{1-\frac{1}{4}}}=\frac{1}{12\pi\sqrt{3}}$

$\therefore (X,Y)\sim BN\left(2,3,9,16,\frac{1}{2}\right)$

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

${\sigma_X}^2 = E\big(X - E(X)\big)^2 = \frac{b^2}{a^2}E\big(Y - E(Y)\big)^2 = \frac{b^2}{a^2}{\sigma_Y}^2$

$$\therefore \rho = \frac{cov(X,Y)}{\sigma_X \cdot \sigma_Y} = \frac{-\frac{b}{a}{\sigma_Y}^2}{\sqrt{{\sigma_Y}^2}\sqrt{\frac{b^2}{a^2}{\sigma_Y}^2}} = \frac{-\frac{b}{a}{\sigma_Y}^2}{\left|\frac{b}{a}\right|{\sigma_Y}^2} = \frac{-\frac{b}{a}}{\left|\frac{b}{a}\right|}$$

$$\therefore \rho = \frac{cov(X,Y)}{\sigma_X \cdot \sigma_Y} = \begin{cases} 1\,, & if\ a\ and\ b\ have\ opposite\ signs \\ -1\,, & if\ a\ and\ b\ have\ same\ signs \end{cases}$$

2.7 .Correlation coefficient and Bivariate Normal Distribution

Exercise:

1. Find the correlation coefficient between X and Y for each of the j.p.d.f. $f(x,y)$ of (X,Y) given below:

(i) $$f(x,y) = \begin{cases} \frac{3}{2}(x^2+y^2) & , \quad 0 \le x \le 1, \ 0 \le y \le 1 \\ 0 & , \quad otherwise \end{cases}$$

(ii) $$f(x,y) = \begin{cases} (x+y) & , \quad 0 \le x, y \le 1 \\ 0 & , \quad otherwise \end{cases}$$

(iii) $$f(x,y) = \begin{cases} 2xy & , \quad 0 < x < 1, \ 0 < y < 1 \\ 0 & , \quad otherwise \end{cases}$$

2. If X, Y and Z are uncorrelated r.vs with 0 mean and standard deviations $5, 12$ and 9 repecitvely and $U = X + Y$ and $V = Y + Z$, then find the correlation coefficient between U and V.

3. If X, Y, Z are uncorrelated r.vs having same variance, find the correlation coefficient between $(X+Y)$ and $(Y+Z)$.

4. If the independent r.vs X and Y have variance 36 and 16 respectively, find the correlation coefficient between $(X+Y)$ and $(X-Y)$.

ANSWERS

1. (i) -0.2055 (ii) $-\frac{1}{11}$ (iii) 0.8

2. $\frac{48}{65}$

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

4. $\frac{5}{13}$

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

$$\boldsymbol{P}[|\boldsymbol{X}-\boldsymbol{\mu}| < k\sigma] \geq \mathbf{1} - \frac{\mathbf{1}}{\boldsymbol{k^2}} \quad \dots\dots\dots\dots(4)$$

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 X and 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,

$$4\big(E(XY)\big)^2 - 4E(X^2)E(Y^2) \leq 0$$

$$\Rightarrow \big(E(XY)\big)^2 \leq E(X^2)E(Y^2)$$

Hence the result.

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

$$\boldsymbol{f(x,y) = \frac{x+y}{21} \text{ for } x = 1, 2, 3 \text{ and } y = 1, 2.}$$

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.

X / Y	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 $\left(E(XY)\right)^2 \leq\ E(X^2)E(Y^2)$.

P1:

A symmetric die is thrown 600 times. Find the lower bound for the probability of getting 80 to 120 sixes.

Solution:

Let X be the total number of sixes.

Then $X \sim B\left(600, \frac{1}{6}\right)$, $E(X) = np = 600 \times \frac{1}{6} = 100$

and $V(X) = np(1-p) = 600 \times \frac{1}{6} \times \frac{5}{6} = \frac{500}{6}$.

Using Chebychev's inequality, we get

$$P\{|X - E(X)| < k\sigma\} \geq 1 - \frac{1}{k^2} \Rightarrow P\left\{|X - 100| < k\sqrt{\frac{500}{6}}\right\} \geq 1 - \frac{1}{k^2}$$

Therefore, $P\left\{100 - k\sqrt{\frac{500}{6}} < X < 100 + k\sqrt{\frac{500}{6}}\right\} \geq 1 - \frac{1}{k^2}$

Taking $k = \frac{20}{\sqrt{\frac{500}{6}}}$, we get $P(80 < X < 120) \geq 1 - \frac{1}{400 \times \left(\frac{6}{500}\right)} = \frac{19}{24}$

The lower bound for the probability of getting 80 to 120 sixes.

P2:

For geometric distribution $p(x) = 2^{-x}$; $x = 1, 2, 3, \ldots$, prove that Chebychev's inequality gives $P\{|X - 2| \le 2\} > \frac{1}{2}$, while the actual probability is $\frac{15}{16}$.

Solution:

$$E(X) = \sum x\, p(x) = \sum_{x=1}^{\infty} \frac{x}{2^x} = \frac{1}{2} + \frac{2}{2^2} + \frac{3}{2^3} + \frac{4}{2^4} + \ldots$$

$$= \frac{1}{2}\left(1 + 2A + 3A^2 + 4A^3 + \ldots\right) = \frac{1}{2}(1 - A)^{-2} = 2, \left(A = \frac{1}{2}\right)$$

$$E(X^2) = \sum x^2\, p(x) = \sum_{x=1}^{\infty} \frac{x^2}{2^x} = \frac{1}{2} + \frac{4}{2^2} + \frac{9}{2^3} + \ldots$$

$$= \frac{1}{2}\{1 + 4A + 9A^2 + \ldots\}\ , where\ A = \frac{1}{2}$$

$$= \frac{1}{2}(1 + A)(1 - A)^{-3} = 6$$

$$\therefore Var(X) = \sigma^2 = E(X^2) - \{E(X)\}^2 = 6 - 4 = 2 \Rightarrow \sigma = \sqrt{2}$$

Using Chebychev's inequality, we get $P\{|X - E(X)| > k\sigma|\} \le \frac{1}{k^2}$

With $k = \sqrt{2}$, we get $P\{|X - 2| > \sqrt{2} \cdot \sqrt{2}\} \le \frac{1}{2} \Rightarrow P[|X - 2| \le 2] > 1 - \frac{1}{2} = \frac{1}{2}$

The actual probability is given by

$$P\{|X - 2| \le 2\} = P\{0 \le X \le 4\} = P\{X = 1,2,3\ or\ 4\}$$

$$= \frac{1}{2} + \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^3 + \left(\frac{1}{2}\right)^4 = \frac{15}{16}$$

P3:

Does there exist a variate $\boldsymbol{X}$ for which

$$\boldsymbol{P[\mu - 2\sigma \leq X \leq \mu + 2\sigma] = 0.6} \quad \textbf{..........} \ \mathbf{(1)}$$

Solution:

We have

$$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = P(|X - \mu| \leq 2\sigma)$$

By Chebychev's inequality $P(|X - \mu| \leq 2\sigma) \geq 1 - \frac{1}{2^2} = 0.75$

Since lower bound for the probability is 0.75, there does not exist a r.v. X for which the equation (1) holds.

P4:

(a) For a p.m.f

$$\boldsymbol{p(x) = \begin{cases} \frac{1}{8} & , \quad x = -1 \\ \frac{6}{8} & , \quad x = 0 \\ \frac{1}{8} & , \quad x = 1 \end{cases}}$$

Find $\boldsymbol{P(|X-\mu| \geq 2\sigma)}$.

(b) Compare this result with that obtained by using Chebychev's inequality.

Solution:

(a)

x:	-1	0	1
$p(x)$:	$\frac{1}{8}$	$\frac{6}{8}$	$\frac{1}{8}$

$\therefore \mu = E(X) = \sum x\, p(x) = -1 \times \frac{1}{8} + 1 \times \frac{1}{8} = 0$ and

$$E(X^2) = \sum x^2 p(x) = 1 \times \frac{1}{8} + 1 \times \frac{1}{8} = \frac{1}{4}$$

$$\therefore Var(X) = E(X^2) - \{E(X)\}^2 = \frac{1}{4} \Longrightarrow \sigma = \frac{1}{2}$$

$$P\{|X-\mu| \geq 2\sigma\} = P\{|X| \geq 1\} = 1 - P(|X| < 1)$$

$$= 1 - P(-1 < X < 1) = 1 - P(X = 0) = \frac{1}{4}$$

(b) $P\{|X-\mu| \geq 2\sigma\} \leq \frac{1}{2^2}$ (By Chebychev's inequality: $P\{|X-\mu| \geq k\sigma\} \leq \frac{1}{k^2}$)

In this case, both results are same.

Note: This example shows that, in general, Chebychev's inequality cannot be improved.

3.1 .Probability inequalities

Exercise:

1. The Chebychev's inequality for random variable X is $(-2 < X < \infty) \geq \frac{21}{25}$, find $E(X)$ and $V(X)$.

2. Two unbiased dice are thrown. If X is the sum of the numbers showing up, prove that $P(|X-7| \geq 3) \leq \frac{35}{54}$. Compare this with the actual probability.

3. If X is the number scored in a throw of a fair die, find the upper bound for $P(|X-\mu| \geq 2.5)$ where $\mu = E(X)$. Also find the actual probability.

4. If X is a r.v. such that $E(X) = 3$ and $E(X^2) = 13$, find the lower bound of $P(-2 \leq X \leq 8)$.

5. A discrete random variable X is specified by $p(-a) = p(a) = \frac{1}{8}$ and $p(0) = \frac{3}{4}$. Compute
   (i) $P(|X| \geq 2\sigma)$ and
   (ii) Chebychev's inequality bound.

Answers:

1. $E(X) = 3$ and $V(X) = 4$
2. Actual probability $= \frac{1}{3}$
3. Upper bound $= 0.47$ and actual probability $= 0$
4. $\frac{21}{25}$
5. (i) $\frac{1}{4}$ (ii) $\frac{1}{4}$

Probability & Statistics
HW 7 - Chebyshev & Empirical & Z score

Name: Key

1. The average IQ of the students in one calculus class is 110, with a standard deviation of 5; the average IQ of students in another class is 106, with a standard deviation of 4. A student has an IQ of 112, in which class is he ranked higher?

$$z_1 = \frac{112-110}{5} \qquad z_2 = \frac{112-106}{4}$$

$$z_1 = 0.4 \qquad z_2 = 1.5$$

$\rightarrow$ $\boxed{\text{Ranked higher in 2nd Class.}}$

2. The average price of the wagon at Car Dealer A is $25,000, with a standard deviation of $4000. The average price at Car Dealer B is $20,000, with a standard deviation of $2000. If I spent $23,000 at dealer A and my sister paid $18,000 at dealer B, which was a better deal?

$$z_A = \frac{23{,}000 - 25{,}000}{4{,}000} \qquad z_B = \frac{18{,}000 - 20{,}000}{2000}$$

$$z_A = -0.5 \qquad z_B = -1$$

$\rightarrow$ $\boxed{\text{Dealer B was the better deal.}}$

3. The average score on an English final examination was 85, with a standard deviation of 5; the average score on a history final exam was 110, with a standard deviation of 8. I made an 80 on the English test and a 100 on the history test. Which test was better?

$$z_E = \frac{80-85}{5} \qquad z_H = \frac{100-110}{8}$$

$$z_E = -1 \qquad z_H = -1.25$$

$\rightarrow$ $\boxed{\text{English}}$

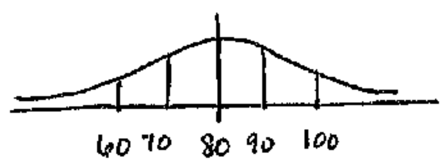
4. The average age of the accountants at Three Rivers Corp. is 26, with a standard deviation of 6; the average salary of the accountants is $31,000 with a standard deviation of $4000. I'm applying at both places and I'm 31 and want to make $35,000, in which category will I be higher?

$$z_{Age} = \frac{31-26}{6} \qquad z_{Salary} = \frac{35000-31000}{4000}$$

$$z_A = 0.83 \qquad z_S = 1$$

$\rightarrow$ $\boxed{\text{Salary}}$

5. Using Chebyshev's theorem, solve the following problems for a distribution with a mean of 80 and a standard deviation of 10



a. At least what percentage of values will fall between 60 and 100?

$$K = 2 \rightarrow \left(1 - \frac{1}{2^2}\right)(100) = \boxed{75\%}$$

b. At least what percentage of values will fall between 65 and 95?

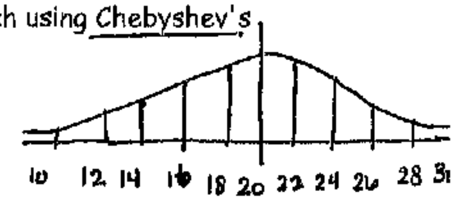
$$K = 1.5 \rightarrow \left(1 - \frac{1}{1.5^2}\right)(100) = \boxed{56\%}$$

$$z = \frac{95-80}{10} \qquad z = \frac{65-80}{10}$$

$$z = 1.5 \qquad z = -1.5$$

Thus, $K = 1.5$

6. The mean of a distribution is 20 and the standard deviation is 2. Answer each using Chebyshev's theorem.



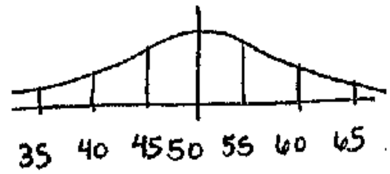
a. At least what percentage of the values will fall between 10 and 30?

K = 5 from Graph or $z = \frac{30-20}{2}$, $z = 5$ $\left[1 - \frac{1}{(5)^2}\right](100) = \boxed{96\%}$

b. At least what percentage of the values will fall between 12 and 28?

K = 4 or $z = \frac{28-20}{2}$, $z = 4$ $\left[1 - \frac{1}{(4)^2}\right](100) = \boxed{93.75\%}$

7. The mean of a distribution is 50 and the standard deviation is 5. Answer each using The Empirical Rule.



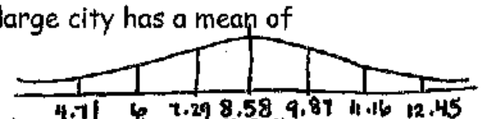
a. At least what percentage of the values will fall between 45 and 55?

1 st dev $\rightarrow$ $\boxed{68\%}$ OR $z = \frac{55-50}{5}$, $z = 1$

b. At least what percentage of the values will fall between 35 and 65?

3 st dev $\rightarrow$ $\boxed{99.7\%}$

8. A sample of the hourly wages of employees who work in restaurants in a large city has a mean of \$8.58 and a standard deviation of \$1.29.



a. Using Chebyshev's theorem, find the range in which at least 75% of the data values will fall.

$.75 = 1 - \frac{1}{K^2}$ $\quad -\frac{1}{K^2} = -0.25 \rightarrow -0.25K^2 = -1$ $\quad K^2 = 4$ $\quad K = 2$ $\quad \therefore\ 8.58 \pm 2(1.29)$ $\boxed{\$6 \text{ to } \$11.16}$

b. Using the Empirical Rule, find the range in which at least 95% of the data values will fall.

95% $\rightarrow$ 2 st dev Thus, range is $\boxed{\$6 \text{ to } \$11.16}$

9. A sample of the labor costs per hour to assemble a certain product has a mean of \$15.72 and a standard deviation of \$2.15. Using Chebyshev's theorem, find the values in which at least 88.89% of the data will lie.

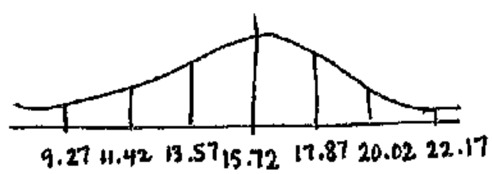
$.8889 = 1 - \frac{1}{K^2}$

$-0.1111 = -\frac{1}{K^2}$

$0.1111K^2 = 1$

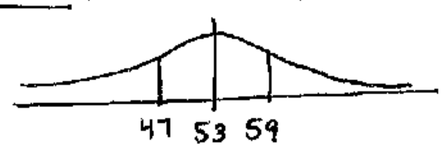
$K^2 = 9 \rightarrow K = 3$

$\therefore\ 15.72 \pm 3(2.15)$ $\boxed{\$9.27 \text{ to } \$22.17}$



10. The average score on a special test of knowledge of wood refinishing has a mean of 53 and a standard deviation of 6. Using the Empirical Rule, find the values in which at least 68% of the data will lie.

68% $\rightarrow$ 1st dev $\quad \therefore\ 53 \pm 1(6)$ $\boxed{47 \text{ to } 59}$



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

$$P(X = x) = \frac{1}{6}, \quad x = 1, 2, 3, 4, 5, 6$$

$$\therefore \quad M_X(t) = \sum_{X=1}^{6} e^{tx} P(X = x) = \frac{1}{6} \sum_{x=1}^{6} e^{tx}$$

$$= \frac{1}{6}(e^t + e^{2t} + e^{3t} + e^{4t} + e^{5t} + e^{6t})$$

$$E(X) = \left[\frac{d}{dt} M_X(t)\right]_{t=0} = \frac{1}{6}[e^t + 2e^{2t} + 3e^{3t} + 4e^{4t} + 5e^{5t} + 6e^{6t}]_{t=0}$$

$$= \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = \frac{21}{6} = \frac{7}{2}$$

$\therefore$ 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,

$$M_X(t) = E(e^{tx}) = \sum_{r=0}^{\infty} e^{tx} P(X = x)$$

$M_X(t) = P(X = 0) + e^t P(X = 1) + e^{2t} P(X = 2) + e^{3t} P(X = 3) +$…+… …… (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,

$$P(X = 0) = 0.4,\ P(X = 1) = 0.6$$

$$P(X = 2) = P(X = 3) = \cdots = 0 \quad \Rightarrow P(X \geq 2) = 0$$

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

$p(x) = \binom{n}{x} p^x q^{n-x},\ x = 0, 1, 2, \dots, n$ and $q = 1 - p$.

Then the m.g.f. of X is given by

$$M_X(t) = E\left[e^{tX}\right] = \sum_{x=0}^{n} e^{tx} p(x) = \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x q^{n-x} = \sum_{x=0}^{n} \binom{n}{x} \left(pe^t\right)^x q^{n-x}$$

$\Rightarrow \ M_X(t) = (q + pe^t)^n$

$\frac{d}{dt} M_X(t) = n(q + pe^t)^{n-1} pe^t \Rightarrow$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

$$p(x) = P(X = x) = \frac{e^{-\lambda}\lambda^x}{\lambda!}\ , x = 0, 1, \ldots \text{ and } \lambda > 0$$

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, $\frac{d}{dt}\left(M_X(t)\right) = p^r(-r)(1 - qe^t)^{-(r+1)}(-qe^t) = qrp^r(1 - qe^t)^{-(r+1)}e^t$

Mean$= \mu = \mu_1' = \left[\frac{d}{dt}\left(M_X(t)\right)\right]_{t=0} = qrp^r(1 - q)^{-(r+1)} = qrp^r(p)^{-(r+1)} = \frac{rq}{p}$

Further, $\frac{d^2}{dt^2}\left(M_X(t)\right) = rqp^r \frac{d}{dt}\left\{(1 - qe^t)^{-(r+1)}e^t\right\}$

$$= rqp^r\left\{-(r+1)(1 - qe^t)^{-(r+2)}(-qe^{2t}) + (1 - qe^t)^{-(r+1)}e^t\right\}$$

Then $\mu_2' = \left[\frac{d^2}{dt^2}\left(M_X(t)\right)\right]_{t=0} = rqp^r\left[(r+1)qp^{-(r+2)} + p^{-(r+1)}\right]$

$$= r(r+1)q^2p^{-2} + rqp^{-1} = \frac{rq}{p}\left(\frac{(r+1)q}{p} + 1\right) = \frac{rq}{p^2}(rq + 1)$$

$$\Rightarrow \quad \mu_2' = \frac{r^2q^2}{p^2} + \frac{rq}{p^2}$$

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 $t > 0$.

3.2

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

$$\boldsymbol{M_X(t) = E(e^{tX})}$$

$$\therefore\ M_X(t) = E(e^{tX}) = E\left[1+\frac{tX}{1!}+\frac{(tX)^2}{2!}+\frac{(tX)^3}{3!}+\cdots\right]$$

$$= E(1)+\frac{t}{1!}E(X)+\frac{t^2}{2!}E(X^2)+\frac{t^3}{3!}E(X^3)+\cdots+\frac{t^r}{r!}E(X^r)+\cdots+\infty$$

$$\therefore\ M_X(t) = 1+\frac{t}{1!}\mu_1'+\frac{t^2}{2!}\mu_2'+\frac{t^3}{3!}\mu_3'+\cdots+\frac{t^r}{r!}\mu_r'+\cdots+\infty \qquad \dots\dots \quad (1)$$

$$M_X(t) = \sum_{r=0}^{\infty}\frac{t^r}{r!}\mu_r'$$

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,

$$\mu_r' = \left[\frac{d^r}{dt^r} M_X(t)\right]_{t=0}, \; r = 1, 2, 3, \ldots$$

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, $i.e.$, $M_{X_1+X_2+X_3+\cdots+X_n}(t) = M_{X_1}(t)M_{X_2}(t)M_{X_3}(t)\ldots M_{X_n}(t)$

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 $P(X = x) = \frac{1}{2^x}, x = 1, 2, 3, \ldots$ 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,

$$M_X(t) = E(e^{tx}) = \sum_{r=0}^{\infty} e^{tx} P(X = x)$$

$M_X(t) = P(X = 0) + e^t P(X = 1) + e^{2t} P(X = 2) + e^{3t} P(X = 3)$ +...+... ...... (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

$$p(x) = P(X = x) = \frac{e^{-\lambda}\lambda^x}{\lambda!}\ , x = 0, 1, \ldots \text{ and } \lambda > 0$$

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, $\frac{d}{dt}\left(M_X(t)\right) = p^r(-r)(1 - qe^t)^{-(r+1)}(-qe^t) = qrp^r(1 - qe^t)^{-(r+1)}e^t$

Mean$= \mu = \mu_1' = \left[\frac{d}{dt}\left(M_X(t)\right)\right]_{t=0} = qrp^r(1 - q)^{-(r+1)} = qrp^r(p)^{-(r+1)} = \frac{rq}{p}$

Further, $\frac{d^2}{dt^2}\left(M_X(t)\right) = rqp^r \frac{d}{dt}\left\{(1 - qe^t)^{-(r+1)}e^t\right\}$

$$= rqp^r\left\{-(r+1)(1 - qe^t)^{-(r+2)}(-qe^{2t}) + (1 - qe^t)^{-(r+1)}e^t\right\}$$

Then $\mu_2' = \left[\frac{d^2}{dt^2}\left(M_X(t)\right)\right]_{t=0} = rqp^r\left[(r+1)qp^{-(r+2)} + p^{-(r+1)}\right]$

$$= r(r+1)q^2p^{-2} + rqp^{-1} = \frac{rq}{p}\left(\frac{(r+1)q}{p} + 1\right) = \frac{rq}{p^2}(rq + 1)$$

$$\Rightarrow \quad \mu_2' = \frac{r^2q^2}{p^2} + \frac{rq}{p^2}$$

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.

$$\boldsymbol{p(x) = \begin{cases} \dfrac{1}{x(x+1)} & , \quad x = 1, 2 \dots \\ 0 & , \quad otherwise \end{cases}}$$

Show that $\boldsymbol{E(X)}$ does not exist even though m.g.f. exist.

Solution:

$$E(X) = \sum_{x=1}^{\infty} x\, p(x) = \sum_{x=1}^{\infty} \frac{1}{x+1} = \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \dots = \sum_{x=1}^{\infty} \frac{1}{x} - 1$$

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 $z = e^t$,

$$M_X(t) = \sum_{x=1}^{\infty} \frac{z^x}{x(x+1)} = \frac{z}{1.2} + \frac{z^2}{2.3} + \frac{z^3}{3.4} + \dots$$

$$= z\left(1 - \frac{1}{2}\right) + z^2\left(\frac{1}{2} - \frac{1}{3}\right) + z^3\left(\frac{1}{3} - \frac{1}{4}\right) + \cdots$$

$$= \left(z + \frac{z^2}{2} + \frac{z^3}{3} + \cdots\right) - \frac{z}{2} - \frac{z^2}{3} - \frac{z^4}{4} \dots$$

$$= -\log(1 - z) - \frac{1}{z}\left(\frac{z^2}{2} + \frac{z^3}{3} + \frac{z^4}{4} + \cdots\right)$$

$$= -\,log(1-z) + 1 + \frac{1}{z}\,log(1-z)\,, |z| < 1$$

$$= 1 + \left(\frac{1}{z} - 1\right) log(1-z)\,,\ \ |z| < 1$$

$$M_X(t) = \begin{cases} 1 + (e^{-\,t} - 1)\,log(1-e^{t}) & , \qquad\qquad t < 0 \\ 1 & , \qquad for \quad t = 0 \end{cases}$$

and $M_X(t)$ does not exist for $t > 0$.

P1:

Find the m.g.f. of uniform distribution $U[a, b]$ and hence obtain the mean and variance of the distribution.

P2:

Find the m.g.f. of Normal $N(\mu, \sigma^2)$ distribution and hence find its mean and variance.

Solution:

Since $X \sim N(\mu, \sigma^2)$, its p.d.f. is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$$

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

$$M_X(t) = E\left[e^{tX}\right] = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{tx} exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$$

Let $z = \frac{x-\mu}{\sigma}$. Thus $x = \mu + \sigma z$ and $dx = \sigma.dz$

Thus, $M_X(t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} exp\left(t(\mu+\sigma z)\right) exp\left(-\frac{z^2}{2}\right) dz$

$$= \frac{1}{\sqrt{2\pi}} e^{t\mu}\int_{-\infty}^{\infty} exp\left[-\frac{1}{2}\left(z^2 - 2t\sigma z\right)\right] dz$$

$$= \frac{1}{\sqrt{2\pi}} e^{t\mu}\int_{-\infty}^{\infty} exp\left[-\frac{1}{2}\left\{(z-\sigma t)^2 - \sigma^2 t^2\right\}\right] dz$$

$$= \frac{1}{\sqrt{2\pi}} e^{t\mu+\frac{1}{2}\sigma^2 t^2}\int_{-\infty}^{\infty} exp\left[-\frac{1}{2}(z-\sigma t)^2\right] dz$$

Let $u = z - \sigma t \quad \Rightarrow du = dz$

$$= e^{t\mu+\frac{1}{2}\sigma^2 t^2} \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}u^2} du$$

$$= e^{t\mu+\frac{1}{2}\sigma^2 t^2}$$

$\Longrightarrow M_X(t) = e^{t\mu+\frac{1}{2}\sigma^2 t^2} \Longrightarrow M_X'(t) = e^{t\mu+\frac{1}{2}\sigma^2 t^2}(\mu + \sigma^2 t)$

$\Longrightarrow \mu$ =Mean= $\mu_1' = M_X'(t)|_{t=0} = \mu$

Further, $M_X''(t) = e^{t\mu+\frac{1}{2}\sigma^2 t^2}(\mu + \sigma^2 t)^2 + e^{t\mu+\frac{1}{2}\sigma^2 t^2}(\sigma^2)$

$\Longrightarrow \mu_2' = M_X''(t)|_{t=0} = \mu^2 + \sigma^2 \Longrightarrow \mu_2' = \mu^2 + \sigma^2$

The variance is given by

Variance $= \mu_2' - (\mu_1')^2 = \mu^2 + \sigma^2 - \mu^2 = \sigma^2.$

P3:

Find the m.g.f. of geometric $G(p)$ distribution and hence obtain its mean and variance.

Solution:

If $X \sim G(p)$, its p.m.f is given by $p(x) = q^x p$ for $x = 0, 1, 2, \ldots, \quad 0 < p < 1.$

Then the m.g.f. is given by

$$M_X(t) = E\left[e^{tX}\right] = \sum_{x=0}^{\infty} e^{tx} q^x p = \sum_{x=0}^{\infty} \left(qe^t\right)^x = \frac{p}{1-qe^t}$$

$$\Rightarrow \; M_X(t) = \frac{p}{1 - qe^t}$$

Then $\mu_1' = M_X'(t)|_{t=0} = pq(1-q)^{-2} = \frac{q}{p}$ and $\mu_2' = M_X''(t)|_{t=0} = \frac{q}{p} + \frac{2q^2}{p^2}$

and $\mu_2 = \mu_2' - (\mu_1')^2 = \frac{q}{p} + \frac{2q^2}{p^2} - \left(\frac{q}{p}\right)^2 = \frac{q}{p} + \frac{q^2}{p^2} = \frac{qp+q^2}{p^2} = \frac{q(p+q)}{p^2} = \frac{q}{p^2}$

Thus, Mean$= \mu = \mu_1' = \frac{q}{p}$ and variance $= \sigma^2 = \frac{q}{p^2}$

P4:

Find the m.g.f. of exponential $E(\lambda)$ distribution and hence find its mean and variance.

Solution:

Since $X \sim E(\lambda)$, its p.d.f. is given by $f(x) = \lambda e^{-\lambda x}$ for $x > 0$ and $\lambda > 0$

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

$$M_X(t) = E\left[e^{tX}\right] = \lambda \int_0^\infty e^{tx} e^{-\lambda x} dx = \lambda \int_0^\infty e^{-(\lambda - t)x} dx$$

$$= \frac{\lambda}{\lambda - t} = \left(1 - \frac{t}{\lambda}\right)^{-1} = \sum_{r=0}^{\infty} \left(\frac{t}{\lambda}\right)^r \quad for\ \lambda > t$$

$$\Rightarrow M_X(t) = \sum_{r=0}^{\infty} \left(\frac{t}{\lambda}\right)^r \quad for\ \lambda > t$$

But $\mu_r' =$ coefficient of $\frac{t^r}{r!}$ in $M_X(t) = \frac{r!}{\lambda^r}$ for $r = 1, 2, \ldots$

Thus, $\mu_1' = \frac{1}{\lambda}$ and $\mu_2' = \frac{2}{\lambda^2}$ and hence $\mu_2 = \mu_2' - (\mu_1')^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$

Thus, $\mu = \frac{1}{\lambda}$ and $\sigma^2 = \frac{1}{\lambda^2}$

3.2. Moment Generating Function

Exercise:

1. Find the m.g.f of a.r.v. whose moments are given by $\mu_r' = (r+1)!\,2^r$

2. If $M_X(t) = \frac{3}{3-t}$, find standard deviation of X

3. Find the m.g.f. of a.r.v X whose p.d.f is given by

$$f(x) = \begin{cases} \frac{x}{2} & , \ 0 \le x \le 2 \\ 0 & , \ otherwise \end{cases}$$

4. Find the m.g.f and hence find the mean and variance of a.r.v. X whose p.d.f. is given by

   i. $$f(x) = \begin{cases} 2e^{-2x} & , \quad x \ge 0 \\ 0 & , \quad otherwise \end{cases}$$

   ii. $$f(x) = \begin{cases} \frac{1}{3} & , \quad -1 < x < 2 \\ 0 & , \quad otherwise \end{cases}$$

   iii. $$f(x) \begin{cases} \frac{1}{3}e^{-\frac{x}{3}} & , \quad x \ge 0 \\ 0 & , \quad otherwise \end{cases}$$

   iv. $$f(x) \begin{cases} \frac{1}{k} & , \quad 0 < x < k \\ 0 & , \quad otherwise \end{cases}$$

   v. $$f(x) = \lambda e^{-\lambda(x-a)}, x \ge a$$

Answers:

1. $M_X(t) = \frac{1}{(1-2t)^r}$

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

3. $\frac{1}{2t^2}(1 + 2t\, e^{2t} - e^{2t})$

4.

i. $M_X(t) = \frac{2}{2-t}\ ,\ \mu = \frac{1}{2}\ ,\ \sigma^2 = \frac{1}{4}$

ii. $M_X(t) = \begin{cases} \frac{e^{2t}-e^{-t}}{3t} & , \quad t \neq 0 \\ 1 & , \quad t = 0 \end{cases}$

iii. $M_X(t) = (1-3t)^{-1},\ \mu = 3\ ,\ \sigma^2 = 9$

iv. $M_X(t) = \frac{e^{tk}-1}{kt}\ ,\mu = \frac{k}{2}\ ,\ \ \sigma^2 = \frac{k^2}{\sqrt{2}}$

v. $M_X(t) = \frac{\lambda e^{at}}{\lambda-t}\ ,\mu = \frac{9\lambda+1}{\lambda}, \sigma^2 = \frac{1}{\lambda^2}$

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. $|\boldsymbol{\emptyset_X(t)}| \leq 1$ 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 $h \neq 0$,

$|\emptyset_X(t+h) - \emptyset_X(t)| = \left|E\left(e^{i(t+h)X}\right) - E\left(e^{itX}\right)\right| = \left|E\left(e^{i(t+h)X} - e^{itX}\right)\right|$

$= \left|E\left\{e^{itX}\left(e^{ihX} - 1\right)\right\}\right| \leq E\left(\left|e^{itX}\right|\left|e^{ihX} - 1\right|\right) = E\left(\left|e^{ihX} - 1\right|\right) \longrightarrow 0$ 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. $\emptyset_X(-t) = \overline{\emptyset_X(t)}$, $i.e$, $\emptyset_X(-t)$ is the complex conjugate of $\emptyset_X(t)$.

Proof: Here $\overline{\emptyset_X(t)} = \overline{E[e^{itX}]} = E[\cos tX - i\sin tX]$

$\Rightarrow \emptyset_X(-t) = E[\cos(-tX) + i\sin(-tX)] = E[\cos tX - i\sin tX] = \overline{\emptyset_X(t)}$

Thus, $\emptyset_X(-t) = \overline{\emptyset_X(t)}$

5. If the p.d.f. is even $i.e.$, $f(-x) = f(x)$, then the characteristic function is real valued and even function of t.

Proof: We know that , $\phi_X(t) = E\left(e^{itX}\right) = \int_{-\infty}^{\infty} e^{itX} f(x)dx$

Let $x = -y \Rightarrow dx = -dy$. Then

$$\phi_X(t) = \int_{\infty}^{-\infty} e^{-ity} f(-y)(-dy) = \int_{-\infty}^{\infty} e^{-ity} f(y)dy \quad (\because\ f \text{ is an even function})$$

$$= E\left[e^{-itX}\right] = \emptyset_X(-t)$$

Thus, $\emptyset_X(t) = \emptyset_X(-t)$

$\Rightarrow \emptyset_X(t)$ 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 $X_1, X_2, \dots, X_n$ are independent, then

$$\boldsymbol{\emptyset_{X_1+\dots+X_n}(t) = \emptyset_{X_1}(t)\,\emptyset_{X_2}(t)\dots\emptyset_{X_n}(t)}$$

Proof:

$$\emptyset_{X_1+\dots+X_n}(t) = E\left[e^{it(X_1+X_2+,\dots+X_n)}\right] = E\left[e^{itX_1}.e^{itX_2}\dots e^{itX_n}\right]$$

$$(\because X_1, X_2, \dots, X_n \text{ are independent})$$

$$= E\left[e^{itX_1}\right].E\left[e^{itX_2}\right]\dots E\left[e^{itX_n}\right]$$

$$= \emptyset_{X_1}(t).\emptyset_{X_2}(t)\dots\emptyset_{X_n}(t)$$

$$\Longrightarrow \emptyset_{X_1+\dots+X_n}(t) = \emptyset_{X_1}(t).\emptyset_{X_2}(t)\dots\emptyset_{X_n}(t)$$

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 $f_1(.)$ and $f_2(.)$ to be identical is that their characteristic function $\emptyset_1(t)$ and $\emptyset_2(t)$ are identical.

Example 1: If $X \sim B(n, p)$, 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

$$\phi_X(t) = E\left[e^{itX}\right] = \sum_{x=0}^{\infty} e^{itx} p(x) = \sum_{x=0}^{\infty} e^{itx} \frac{e^{-\lambda}\lambda^x}{x!}$$

$$= e^{-\lambda} \sum_{x=0}^{\infty} \frac{\left(\lambda e^{it}\right)^x}{x!} = e^{-\lambda} e^{\lambda e^{it}} = e^{\lambda\left(e^{it}-1\right)}$$

$$\Rightarrow j_X(t) = e^{\lambda\left(e^{it}-1\right)}$$

Now, $\frac{d}{dt}\left(\emptyset_X(t)\right) = e^{\lambda(e^{it}-1)}\lambda i e^{it}$

Thus, the mean is given by

$$\mu = \mu_1' = E(X) = (-i)\frac{d}{dt}\left(\phi_X(t)\right)\Big|_{t=0} = (-i)(\lambda i) = \lambda$$

$$\Longrightarrow \ \mu = \lambda$$

Now, $\frac{d^2}{dt^2}\left(\emptyset_X(t)\right) = (\lambda i)\frac{d}{dt}\left[e^{\lambda(e^{it}-1)}e^{it}\right]$

$$= (\lambda i)\left[e^{\lambda(e^{it}-1)}\lambda i e^{2it} + e^{\lambda(e^{it}-1)} i e^{it}\right]$$

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 $\frac{x-\mu}{\sigma} = z \Longrightarrow x = \mu + \sigma z \Longrightarrow dx = \sigma\, dz$

$$= \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{itx} exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right]dx = \int_{-\infty}^{\infty} e^{it(\mu+\sigma z)} e^{-\frac{1}{2}z^2} dz$$

$$= \frac{e^{it\mu}}{\sqrt{2\pi}}\int_{-\infty}^{\infty} exp\left[-\frac{1}{2}\left(z^2 - 2i\sigma zt\right)\right]dz$$

$$= \frac{e^{it\mu}}{\sqrt{2\pi}}\int_{-\infty}^{\infty} exp\left[-\frac{1}{2}\left(z^2 - 2i\sigma zt + i^2\sigma^2t^2 - i^2\sigma^2t^2\right)\right]dz$$

$$= \frac{e^{it\mu - \frac{\sigma^2t^2}{2}}}{\sqrt{2\pi}}\int_{-\infty}^{\infty} exp\left[-\frac{1}{2}(z - i\sigma t)^2\right]dz$$

Let $z - i\sigma t = u \Rightarrow dz = du$

$$= e^{it\mu - \frac{\sigma^2 t^2}{2}} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{u^2}{2}} du$$

$$= e^{it\mu - \frac{\sigma^2 t^2}{2}} \quad \left(\because \quad \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{u^2}{2}} du = 1 \right)$$

$\Rightarrow \ \emptyset_X(t) = e^{itu - \frac{\sigma^2 t^2}{2}}$

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 $\emptyset_X(t) = (q + pe^{it})^n$.

Solution: We have, $\emptyset_X(t) = (q + pe^{it})^n$ and

$$\emptyset_X(t) = (q + pe^{it})^n = \sum_{j=0}^{n} \binom{n}{j} (pe^{it})^j q^{n-j}$$

$$= \sum_{j=0}^{n} \binom{n}{j} p^j q^{n-j} e^{itj} = \sum_{j=0}^{n} P(X = j) e^{itj}$$

$$= E[e^{itX}] \text{ where } P(X = j) = \binom{n}{j} p^j q^{n-j}$$

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, $\emptyset_X(t) = e^{\lambda(e^{it}-1)}$

$$\emptyset_X(t) = e^{\lambda(e^{it}-1)} = e^{-\lambda} e^{\lambda e^{it}} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^{it})^x}{x!} = \sum_{x=0}^{\infty} \frac{e^{-\lambda}\lambda^x}{x!} e^{itx}$$

$$= \sum_{x=0}^{\infty} P(X = x) e^{itx} = E[e^{itX}]$$

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

$$\emptyset(\boldsymbol{t}) = \begin{cases} \boldsymbol{1 - |t|} & , \quad |\boldsymbol{t}| \leq \boldsymbol{1} \\ \boldsymbol{0} & , \quad |\boldsymbol{t}| > 1 \end{cases}$$

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,

$$\int_{0}^{1} e^{-itx}(1-t)dt = \frac{1}{ix} + \frac{1}{(ix)^2}\left(e^{-ix} - 1\right)$$

$$\therefore f(x) = \frac{1}{2\pi}\left[\frac{1}{(ix)^2}\left\{e^{ix} - 1 + e^{-ix} - 1\right\}\right] = \frac{1}{\pi x^2}\left(1 - \frac{e^{ix} + e^{-ix}}{2}\right)$$

$$\Longrightarrow f(x) = \frac{1}{\pi x^2}(1 - cos\, x)\, , -\infty < x < \infty$$

P1:

If $X \sim NB(r, p)$, find its characteristic function 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

$$\emptyset_X(t) = E[e^{itX}] = \sum_{x=0}^{\infty} e^{itx} p(x) = \sum_{x=0}^{\infty} e^{itx} \binom{-r}{x} p^r (-q)^x$$

$$= p^r \sum_{x=0}^{\infty} \binom{-r}{x} \left(-qe^{it}\right)^x = p^r \left(1 - qe^{it}\right)^{-r}$$

Thus, $\emptyset_X(t) = p^r \left(1 - qe^{it}\right)^{-r}$

Now,

$$\frac{d}{dt}\left(\emptyset_X(t)\right) = p^r(-r)\left(1 - qe^{it}\right)^{-(r+1)}(-q)ie^{it} = ip^r qre^{it}\left(1 - qe^{it}\right)^{-(r+1)}$$

Thus, mean $= \mu = \mu_1{}' = (-i)\frac{d}{dt}\left(\emptyset_X(t)\right)\Big|_{t=0}$

$$= (-i)ip^r qr(1 - q)^{-(r+1)} = p^r qr(p)^{-(r+1)} = \frac{qr}{p}$$

$$\Longrightarrow \mu = \frac{rq}{p}$$

Now, $\frac{d^2}{dt^2}\left(\emptyset_X(t)\right) = p^r rqi \frac{d}{dt}\left[\left(1 - qe^{it}\right)^{-(r+1)} e^{it}\right]$

$$= p^r rqi\left[-(r + 1)\left(1 - qe^{it}\right)^{-(r+2)}(-qi)e^{2it} + \left(1 - qe^{it}\right)^{-(r+1)} ie^{it}\right]$$

Thus, $\mu_2{}' = (-i)^2 \frac{d^2}{dt^2}\left(\emptyset_X(t)\right)\Big|_{t=0}$

$$= (-i)^2 p^r rqi\left[(r+1)qip^{-(r+2)} + p^{-(r+1)}i\right]$$

$$= (-i)^2 p^r rqi^2 p^{-(r+2)}[(r+1)q + p]$$

$$= \frac{rq}{p^2}(rq + q + p) = \frac{rq}{p^2}(rq + 1)$$

$$\Longrightarrow \mu_2{}' = \frac{r^2q^2}{p^2} + \frac{rq}{p^2}$$

Thus, the variance is given by

$$\sigma^2 = \mu_2{}' - (\mu_1{}')^2 = \frac{r^2q^2}{p^2} + \frac{rq}{p^2} - \frac{r^2q^2}{p^2}$$

$$\Longrightarrow \sigma^2 = \frac{rq}{p^2}$$

P2:

Find the characteristic function of uniform $\boldsymbol{U[a,b]}$ distribution and hence obtain the mean and variance of the distribution.

Solution:

Since $X \sim U(a,b)$, its p.d.f, is given by

$$f(x) = \frac{1}{b-a}, a < x < b$$

The characteristic function is given by

$$\emptyset_X(t) = E\left[e^{itX}\right] = \int_a^b e^{itx} f(x)dx = \int_a^b e^{itx} \frac{1}{b-a} dx = \frac{1}{b-a} \int_a^b e^{itx} dx$$

$$= \frac{1}{b-a}\left[\frac{e^{itx}}{it}\right]_a^b = \frac{1}{(b-a)it}\left(e^{itb} - e^{ita}\right)$$

$$\Rightarrow \emptyset_X(t) = \frac{e^{itb} - e^{ita}}{it(b-a)}$$

$$\Rightarrow t\emptyset_X(t) = \frac{1}{i(b-a)}\left(e^{itb} - e^{ita}\right) \quad \dots\dots.. \quad (1)$$

Differentiating both sides of (1) w.r.t. t, we get

$$t\emptyset'_X(t) + \emptyset_X(t) = \frac{1}{i(b-a)}\left[(ib)e^{itb} - (ia)e^{ita}\right] \quad \dots\dots \quad (2)$$

Again differentiating both sides of (2) w.r.t. t, we get

$$t\emptyset''_X(t) + \emptyset'_X(t) + \emptyset'_X(t) = \frac{1}{i(b-a)}\left[(ib)^2 e^{itb} - (ia)^2 e^{ita}\right]$$

$$\Rightarrow t\emptyset''_X(t) + 2\emptyset'_X(t) = \frac{1}{i(b-a)}\left[(ib)^2 e^{itb} - (ia)^2 e^{ita}\right]$$

In general,

$$t\frac{d^{k+1}}{dt^{k+1}}\left(\emptyset_X(t)\right) + (k+1)\frac{d^k}{dt^k}\left(\emptyset_X(t)\right) = \frac{1}{i(b-a)}\left[(ib)^{k+1}e^{itb} - (ia)^{k+1}e^{ita}\right]$$

If $t = 0$, then

$$(k+1)\frac{d^k}{dt^k}\left(\emptyset_X(t)\right)\Big|_{t=0} = \frac{1}{i(b-a)}[(ib)^{k+1} - (ia)^{k+1}]$$

$$\Longrightarrow \frac{d^k}{dt^k}\left(\emptyset_X(t)\right)\Big|_{t=0} = \frac{1}{k+1}\frac{1}{i(b-a)}[(ib)^{k+1} - (ia)^{k+1}] \quad \dots\dots\dots\dots\dots \ (3)$$

From (3), the mean is given by

$$\mu = \text{mean} = \mu_1{}' = (-i)\frac{d}{dt}\left(\emptyset_X(t)\right)\Big|_{t=0}$$

$$= (-i)\frac{1}{2i(b-a)}[(ib)^2 - (ia)^2] = \frac{1}{2(b-a)}(b^2 - a^2)$$

$$= \frac{1}{2(b-a)}(b-a)(b+a) = \frac{b+a}{2}$$

$$\Longrightarrow \mu = \frac{b+a}{2}$$

Again from (3),

$$\mu_2{}' = (-i)^2\frac{d^2}{dt^2}\left(\emptyset_X(t)\right)\Big|_{t=0} = (-i)^2\frac{1}{3i(b-a)}((ib)^3 - (ia)^3) =$$

$$= \frac{1}{3(b-a)}(b-a)(b^2 + ab + a^2)$$

$$\Longrightarrow \mu_2{}' = \frac{b^2 + ab + a^2}{3}$$

Thus, the variance is given by

$$\sigma^2 = \mu_2{}' - (\mu_1{}')^2 = \frac{b^2+ab+a^2}{3} - \frac{(b+a)^2}{4} = \frac{4(b^2+ab+a^2) - 3(b^2+2ab+a^2)}{12} = \frac{(b-a)^2}{12}$$

$$\Longrightarrow \sigma^2 = \frac{(b-a)^2}{12}$$

P3:

Find the characteristic function of exponential $E(\lambda)$ distribution and hence find its mean and variance.

Solution:

Since $X \sim E(\lambda)$, its p.d.f. is given by

$$f(x) = \lambda e^{-\lambda x}, x > 0, \lambda > 0$$

The characteristic function of X is given by

$$\emptyset_X(t) = E\left[e^{itX}\right] = \int_0^\infty e^{itx} f(x)dx = \lambda \int_0^\infty e^{itx} e^{-\lambda x} dx = \lambda \int_0^\infty e^{-(\lambda - it)x} dx$$

$$= \lambda \left[\frac{e^{-(\lambda-it)x}}{-(\lambda - it)}\right]_0^\infty = \frac{\lambda}{\lambda - it} = \lambda(\lambda - it)^{-1} = \left(1 - \frac{it}{\lambda}\right)^{-1}$$

$$= \sum_{r=0}^{\infty} \left(\frac{it}{\lambda}\right)^r = \sum_{r=0}^{\infty} \frac{(it)^r}{r!} \frac{r!}{\lambda^r}$$

$$\Longrightarrow \emptyset_X(t) = \sum_{r=0}^{\infty} \frac{(it)^r}{r!} \frac{r!}{\lambda^r}$$

Thus, $\mu_r{}' = coef.$ of $\frac{(it)^r}{r!}$ in $\emptyset_X(t) = \frac{r!}{\lambda^r}$.

Hence, the mean is given by $\mu = \mu_1{}' = \frac{1}{\lambda}$

Further, $\mu_2{}' = \frac{2!}{\lambda^2} = \frac{2}{\lambda^2}$ and variance is given by

$$\sigma^2 = \mu_2{}' - (\mu_1{}')^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

Thus, $\sigma^2 = \frac{1}{\lambda^2}$

P4:

Find the p.m.f. of r.v. whose characteristic function is given by

$$\boldsymbol{\phi_X(t) = \frac{p}{1-qe^{it}}}$$

Solution:

Given $\emptyset_X(t) = \frac{p}{1-qe^{it}} = p\left(1 - qe^{it}\right)^{-1}$

$$= p\sum_{x=0}^{\infty}\left(qe^{it}\right)^x = \sum_{x=0}^{\infty} pq^x e^{itx}$$

$$= \sum_{x=0}^{\infty} P(X = x)e^{itx} = E\left[e^{itX}\right]$$

where $P(X = x) = pq^x$, $x = 0,1,2, \ldots$ which is geometric distribution.

3.3. Characteristic Function

Exercise

1. Find the characteristic function of a r.v. X whose moments are given by $\mu_r{}' = (r+1)!\,2^r$.

2. If $\emptyset_X(t) = \frac{3}{3-it}$, then find standard deviation of X.

3. Find the characteristic function of a r.v.X, whose p.d.f. is given by
$$f(x) = \begin{cases} \frac{x}{2} & , \quad 0 \le x \le 2 \\ 0 & , \quad otherwise \end{cases}$$

4. Find the characteristic function and hence obtain the mean and variance of a r.v.X, whose p.d.f. is given by
$$f(x) = \begin{cases} 2e^{-2x} & , \quad x \ge 0 \\ 0 & , \quad otherwise \end{cases}$$

5. Find the characteristic function of a r.v. X whose p.d.f. is given by
$$f(x) = \begin{cases} \frac{1}{3} & , \quad -1 < x < 2 \\ 0 & , \quad otherwise \end{cases}.$$

6. Find the characteristic function and hence find the mean and variance of a r.v.X whose p.d.f. is given by
$$f(x) = \begin{cases} \frac{1}{3}e^{-\frac{x}{3}} & , \quad x \ge 0 \\ 0 & , \quad otherwise \end{cases}$$

7. A r.v.X has the p.d.f. $f(x) = \begin{cases} \frac{1}{k} & , \quad 0 < x < k \\ 0 & , \quad otherwise \end{cases}$.

   Find the characteristic function, mean and variance of X.

8. Find the characteristic function and hence find the mean and variance of a r.v.X whose p.d.f. is given by $f(x) = \lambda e^{-\lambda(x-a)}$, $x \ge a$.

9. Find the p.m.f. whose characteristic function is given by $\emptyset_X(t) = \frac{e^{it}}{2-e^{it}}$.

Answers:

1. $\emptyset_X(t) = \frac{1}{(1-2it)^2}$
2. $\frac{1}{3}$
3. $\frac{1}{2t^2}\left(e^{2it} - 2ite^{2it} - 1\right)$
4. $\emptyset_X(t) = \frac{2}{2-it}$, $\mu = \frac{1}{2}$, $\sigma^2 = \frac{1}{4}$
5. $\emptyset_X(t) = \begin{cases} \frac{e^{2it}-e^{it}}{3it} & , \quad t \neq 0 \\ 1 & , \quad t = 0 \end{cases}$
6. $(1 - 3it)^{-1}$, $\mu = 3, \sigma^2 = 9$
7. $\emptyset_X(t) = \frac{e^{itk}-1}{kit}$, $\mu = \frac{k}{2}, \sigma^2 = \frac{k^2}{12}$
8. $\emptyset_X(t) = \frac{\lambda e^{ait}}{\lambda-it}$, $\mu = \frac{a\lambda+1}{\lambda}, \sigma^2 = \frac{1}{\lambda^2}$
9. $p(x) = \left(\frac{1}{2}\right)^k$, $k = 1,2,3, \ldots$

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

$$k_j = (-i)^j \frac{d^j K_X(t)}{dt^j}\bigg|_{t=0}$$

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

$$K_X(t) = ln\left(\emptyset_X(t)\right) = ln\left[e^{\lambda\left(e^{it}-1\right)}\right] = \lambda\left(e^{it} - 1\right)$$

$$= \lambda\left[1 + (it) + \frac{(it)^2}{2!} + \frac{(it)^3}{3!} + \cdots - 1\right]$$

$$\Rightarrow K_X(t) = (it)\lambda + \frac{(it)^2}{2!}\lambda + \cdots$$

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

$$\emptyset_X(t) = E\left[e^{itX}\right] = \sum_{x=0}^{\infty} e^{itx} p(x) = \sum_{x=0}^{\infty} e^{itx} \binom{-r}{x} p^r (-q)^x$$

$$= p^r \sum_{x=0}^{\infty} \binom{-r}{x} \left(-qe^{it}\right)^x = p^r \left(1 - qe^{it}\right)^{-r}$$

$$\Rightarrow \emptyset_X(t) = p^r\left(1 - qe^{it}\right)^{-r}$$

The c.g.f. is given by

$$K_X(t) = ln(\emptyset_X(t)) = ln\left(p^r(1 - qe^{it})^{-r}\right)$$

$$= r\, ln\, p - r\, ln(1 - qe^{it})$$

Now, $\frac{d}{dt}(K_X(t)) = (-r)\frac{-iqe^{it}}{1-qe^{it}} = \frac{irqe^{it}}{1-qe^{it}}$

$$\therefore k_1 = (-i)\frac{d}{dt}(K_X(t))\Big|_{t=0} = (-i)\frac{(irq)}{1-q} = \frac{rq}{p}$$

And $\frac{d^2}{dt^2}(K_X(t)) = irq\frac{d}{dt}\left[\frac{e^{it}}{1-qe^{it}}\right] = irq\left[\frac{(1-qe^{it})ie^{it}+e^{it}qie^{it}}{(1-qe^{it})^2}\right]$

$$\therefore k_2 = (-i)^2\frac{d^2}{dt^2}(K_X(t))\Big|_{t=0} = (-i)^2(irq)(i)\left[\frac{p+q}{p^2}\right]$$

$$\Longrightarrow k_2 = \frac{rq}{p^2}$$

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

$$K_X(t) = ln\left[e^{it\mu - \frac{1}{2}\sigma^2 t^2}\right] = it\mu - \frac{1}{2}\sigma^2 t^2$$

$$\Longrightarrow K_X(t) = (it)\mu + \frac{(it)^2}{2!}\sigma^2$$

$\therefore k_1 =$ 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 $\boldsymbol{X_1, X_2, \dots, X_m}$ 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 $\boldsymbol{\mu_j}' = \boldsymbol{E}(\boldsymbol{X^j})$ be the $\boldsymbol{j}^{\text{th}}$ moment of $\boldsymbol{X}$ about zero for $\boldsymbol{j} = \mathbf{1}, \mathbf{2}, \mathbf{3}, \dots, \boldsymbol{n}$ where $\boldsymbol{\mu_0}' = \mathbf{1}$. Let $\boldsymbol{k_1}, \boldsymbol{k_2}, \dots, \boldsymbol{k_n}$ be the $\boldsymbol{n}$ cumulants of $\boldsymbol{X}$. Then

$$\mu_{r+1}' = \sum_{j=0}^{r} \binom{r}{j} \mu_j' k_{(r+1-j)} \qquad \text{.... (1)}$$

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

$$\emptyset_X'(t) = e^{K_X(t)} K_X'(t) \qquad \text{...(2)}$$

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

$$\left.\begin{aligned} \mu_1' &= k_1 \\ \mu_2' &= k_2 + \mu_1' k_1 \\ \mu_3' &= k_3 + 2\mu_1' k_2 + \mu_2' k_1 \\ \mu_4' &= k_4 + 3\mu_1' k_3 + 3\mu_2' k_2 + \mu_3' k_1 \end{aligned}\right\} \qquad \text{.....(3)}$$

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** $= \boldsymbol{\mu_2} = \boldsymbol{\sigma^2} = \boldsymbol{k_2}$**.**

P1:

Let X be a.r.v with p.d.f

$$\boldsymbol{f(x)} = \begin{cases} \frac{\mathbf{1}}{\mathbf{2a}}, & -\boldsymbol{a} < x < a \\ \mathbf{0}, & \boldsymbol{otherwise} \end{cases}$$

Find the $c.g.f$.of X and hence find its mean and variance.

Solution:

The characteristic function of X is given by

$$\phi_X(t) = E\left[e^{itX}\right] = \int_{-a}^{a} e^{itx} f(x)dx = \frac{1}{2a}\int_{-a}^{a} e^{itx}dx$$

$$= \frac{1}{2a}\left[\frac{e^{itx}}{it}\right]_{-a}^{a} = \frac{1}{2ait}\left[e^{ita} - e^{-ita}\right]$$

$$= \frac{1}{2ait}\left[\left(1 + ita + \frac{(ita)^2}{2!} + \frac{(ita)^3}{3!} + \cdots\right) - \left(1 - ita + \frac{(ita)^2}{2!} - \frac{(ita)^3}{3!} + \cdots\right)\right]$$

$$= \frac{1}{2ait}\left[2ita + 2\frac{(ita)^3}{3!} + 2\frac{(ita)^5}{5!} + \cdots\right]$$

$$\Rightarrow \emptyset_X(t) = \left[1 + \frac{(it)^2}{2!}\frac{a^2}{3} + \frac{(it)^4}{4!}\frac{a^4}{5} + \cdots\right]$$

The c.g.f. of X is given by

$$K_X(t) = ln[\emptyset_X(t)] = ln\left[1 + \frac{(it)^2}{2!}\frac{a^2}{3} + \frac{(it)^4}{4!}\frac{a^4}{5} + \cdots\right]$$

$$\Rightarrow K_X(t) = \left(\frac{(it)^2}{2!}\frac{a^2}{3} + \frac{(it)^4}{4!}\frac{a^4}{5} + \cdots\right) - \frac{1}{2}\left[\frac{(it)^2}{2!}\frac{a^3}{3} + \frac{(it)^4}{4!}\frac{a^4}{5} + \cdots\right]^2 + \cdots$$

$\Rightarrow k_1$ =coeff. of (it) in $K_X(t) = 0$ and k_2 =coeff.of $\frac{(it)^2}{2!}$ in $K_X(t) = \frac{a^3}{3}$

Thus, the mean and variance are given by $\mu = 0$ and $\sigma^2 = \frac{a^3}{3}$

P2:

Let X be r.v. with p.d.f

$$f(x) = \frac{1}{2} e^{-|x|}, -\infty < x < \infty$$

Find the c.g.f. of and hence find its mean and variance.

Solution:

The characteristic function of X is given by

$$\phi_X(t) = E\left[e^{itX}\right] = \int_{-\infty}^{\infty} e^{itx} f(x)\,dx = \frac{1}{2}\int_{-\infty}^{\infty} e^{itx} e^{-|x|} dx$$

$$= \frac{1}{2}\left[\int_{-\infty}^{\infty} (\cos tx + i \sin tx) e^{-|x|} dx\right] = \frac{1}{2} 2 \int_{0}^{\infty} (\cos t\,x) e^{-|x|} dx$$

(Since the integrals in the first and second integrals are even and odd functions of x repsectively).

$$= \int_0^{\infty} e^{-x} \cos tx\, dx$$

$$= \frac{1}{t^2} - \frac{1}{t^2}\int_0^{\infty} e^{-x} \cos tx\, dx$$

(On integration by parts)

$$\Rightarrow \quad \emptyset_X(t) = \frac{1}{t^2} - \frac{1}{t^2}\, \emptyset_X(t) \Rightarrow \quad \emptyset_X(t) = \frac{1}{1+t^2}$$

The c.g.f. is given by $K_X(t) = ln\big(\emptyset_X(t)\big) = ln\left[\frac{1}{1+t^2}\right] = -\, ln(1+t^2)$

$$= -\left(t^2 - \frac{t^4}{2} + \frac{t^6}{3} - \cdots\right) = -t^2 + \frac{t^4}{2} - \frac{t^6}{3} + \cdots$$

$$\Rightarrow K_X(t) = \frac{(it)^2}{2!} 2 + \cdots$$

$\Rightarrow k_1 = coeff.of\,(it)$ in $K_X(t) = 0$ and $k_2 = coeff.of\ \frac{(it)^2}{2!}$ in $K_X(t) = 2$

Thus, the mean and variance are given by $\mu = 0$ and $\sigma^2 = 2$

P3:

Let X be a.r.v. with p.m.f given by

$$\boldsymbol{p(x)} = \begin{cases} \boldsymbol{q^x p} & , \quad \boldsymbol{x = 0, 1, 2, \ldots,}\ \boldsymbol{0} < p < 1,\ \boldsymbol{q = 1 - p} \\ \boldsymbol{0} & , \quad \boldsymbol{otherwise} \end{cases}$$

Find the c.g.f. and hence find its mean and variance.

Solution:

The characteristic function of X is given by

$$\phi_X(t) = E\left[e^{itX}\right] = \sum_{x=0}^{\infty} e^{itx} p(x) = \sum_{x=0}^{\infty} e^{itx} q^x p$$

$$= p\sum_{x=0}^{\infty}\left(qe^{it}\right)^x = p\left(1 - qe^{it}\right)^{-1}$$

$$\Rightarrow \emptyset_X(t) = p\left(1 - qe^{it}\right)^{-1}$$

The c.g.f. is given by $K_X(t) = ln[\emptyset_X(t)] = ln\, p - ln\left(1 - qe^{it}\right)$

Now, $\frac{d}{dt}\left(K_X(t)\right) = \frac{qie^{it}}{1-qe^{it}} = (iq)\left[\frac{e^{it}}{1-qe^{it}}\right]$

$$\therefore\ k_1 = (-i)\frac{d}{dt}\left(K_X(t)\right)\Big|_{t=0} = (-i)(iq)\frac{1}{p} = \frac{q}{p}$$

Further, $\frac{d^2}{dt^2}\left(K_X(t)\right) = (iq)\left[\frac{(1-qe^{it})ie^{it} + e^{2it}\, iq}{\left(1-qe^{it}\right)^2}\right]$ and

$$k_2 = (-i)^2\frac{d^2}{dt^2}\left(K_X(t)\right)\Big|_{t=0} = (-i)^2(iq)\frac{(ip + iq)}{p^2} \Rightarrow k_2 = \frac{q}{p^2}$$

Thus, the mean and variance are given by $\mu = \frac{q}{p}$ and $\sigma^2 = \frac{q}{p^2}$ respectively.

P4:

Let X be a r.v. with p.d.f.

$$\boldsymbol{f(x) = \frac{1}{2\lambda} exp\left[\frac{-|x-\mu|}{\lambda}\right]},\ -\infty < x, \mu < \infty, \quad \sigma > 0$$

Find the c.g.f. and hence obtain its mean and variance.

Solution:

It can be shown that the characteristic function is given by

$$\emptyset_X(t) = \frac{e^{it\mu}}{1+\lambda^2 t^2} \qquad \text{(do it!)}$$

The c.g.f. is given by $K_X(t) = ln[\emptyset_X(t)] = ln\left[\frac{e^{it\mu}}{1+\lambda^2 t^2}\right]$

$$= it\mu - ln(1 + \lambda^2 t^2)$$

$$= it\mu - (\lambda^2 t^2 + \cdots)$$

$$\Longrightarrow K_X(t) = it\mu + \frac{(it)^2}{2!} 2\lambda^2 + \cdots$$

$\Longrightarrow k_1 = coef.of(it) = \mu$ and $k_2 = coef.of \frac{(it)^2}{2!} = 2\lambda^2$

Thus, the mean and variance are given by μ and $2\lambda^2$ respectively.

3.4. Cumulant Generating Function

Exercise

1. Let $U = \frac{X-a}{h}$ where a and h are real constants. Obtain the cumulants of U interms of cumulants of X.

2. For a distribution, the cumulants are given by
$$k_r = n(r-1)!\,, n > 0$$
Find the ch.f.

3. The p.m.f of X is given by
$$p(x) = q^{r-1}p\,, r = 1, 2, 3, \dots$$
Find the c.g.f. and hence find its mean and variance.

4. The p.d.f. of X is given by
$$f(x) = \frac{e^{-x}}{(1+e^{-x})^2}\,, \qquad -\infty < x < \infty$$
Find the c.g.f. and hence find its mean and variance.

Answers

1. $k_1(U) = \frac{k_1(X) - a}{h}$ and $k_r(U) = \frac{k_r(X)}{h^r}$ for $r = 2, 3 \dots$
2. $\emptyset_X(t) = (1 - it)^{-n}$
3. $\mu = \frac{1}{p}$ and $\sigma^2 = \frac{1+q}{p^2}$
4. $\mu = 0$ and $\sigma^2 = \frac{\pi^2}{3}$

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 X be a discrete random variable with p.g.f. $G_X(t)$. Let $Y = a + bX$ where a and b are real constants. Then $G_Y(t) = t^a G_X(t^b)$

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

$$\boldsymbol{\mu = E(X) = G_X^{(1)}(1)} \text{ **and** }$$

$$\boldsymbol{\sigma^2 = V(X) = G_X^{(2)}(1) + G_X^{(1)}(1) - \left[G_X^{(1)}(1)\right]^2}$$

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

$$\Rightarrow \sigma^2 = G_X^{(2)}(1) + G_X^{(1)}(1) - \left(G_X^{(1)}(1)\right)^2$$

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 ,

$$G_X(t) = \sum_{x=0}^{\infty} t^x p_1(x) \quad and \quad G_Y(t) = \sum_{y=0}^{\infty} t^y p_2(y)$$

Now, $\quad G_{X+Y}(t) = G_X(t)G_Y(t) \quad$ (Since X and Y are independt)

$$= \left(\sum_{x=0}^{\infty} t^x p_1(x)\right)\left(\sum_{y=0}^{\infty} t^y p_2(y)\right)$$

$$\Rightarrow G_{X+Y}(t) = \sum_{x=0}^{\infty}\sum_{y=0}^{\infty} p_1(x)p_2(y)t^{x+y} \qquad \ldots (1)$$

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

$$\sum_{z=0}^{\infty} t^z p(z) = \sum_{x=0}^{\infty}\sum_{y=0}^{\infty} p_1(x)p_2(y)t^{x+y}$$

$$\Longrightarrow \sum_{z=0}^{\infty} t^z p(z) = \sum_{z=0}^{\infty}\left(\sum_{x=0}^{z} p_1(x)p_2(z-x)\right)t^z$$

$$\Longrightarrow p(z) = \sum_{x=0}^{z} p_1(x)p_2(z-x) \text{ , for } z = 0,1,2,\dots$$

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

$$G_X(t) = E[t^X] = \sum_{x=0}^{n} t^x p(x) = \sum_{x=0}^{n} t^x \binom{n}{x} p^x q^{n-x}$$

$$= \sum_{x=0}^{n} \binom{n}{x} (tp)^x q^{n-x} = (q + tp)^n$$

$$\Longrightarrow G_X(t) = (q + tp)^n$$

Differentiating both sides w.r.t., t we get

$$G_X^{(1)}(t) = n(q + tp)^{n-1} p$$

$\Longrightarrow \mu = \text{mean} = \mu_1{}' = G_X^{(1)}(1) = np$ and variance is given by

$$\sigma^2 = G_X^{(2)}(1) + G_X^{(1)}(1) - \left[G_X^{(1)}(1)\right]^2$$

But $G_X^{(2)}(t) = np(n-1)(q + tp)^{n-1} p$

$$\Longrightarrow G_X^{(2)}(t) = n(n-1)p^2 = n^2p^2 - np^2$$

Therefore, $\sigma^2 = n^2p^2 - np^2 + np - n^2p^2 = np(1-p) \Longrightarrow \sigma^2 = npq$

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

$$p(x) = \frac{e^{-\lambda}\lambda^x}{x!},\ x = 0,1,\ldots \text{ and } \lambda > 0$$

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

$$G_X(t) = E[t^X] = \sum_{x=0}^{\infty} t^x p(x) = \sum_{x=0}^{\infty} t^x \frac{e^{-\lambda}\lambda^x}{x!} = e^{-\lambda}\sum_{x=0}^{\infty}\frac{(t\lambda)^x}{x!} = e^{-\lambda}e^{t\lambda} = e^{\lambda(t-1)}$$

$\Rightarrow G_X(t) = e^{\lambda(t-1)}$

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

$$p(x) = \binom{-r}{x} p^x(-q)^x, x = 0,1,2,\ldots$$

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

$$p(x) = q^x p, \qquad x = 0,1,2,\ldots$$

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

$$G_X(t) = E[t^X] = \sum_{x=0}^{\infty} t^x p(x) = \sum_{x=0}^{\infty} t^x q^x p = p\sum_{x=0}^{\infty} (tq)^x = \frac{p}{1-tq}$$

$\Longrightarrow G_X(t) = p(1-tq)^{-1}$

$G_X^{(1)}(t) = p(-1)(1-tq)^{-2}(-q) = pq(1-tq)^{-2}$

$$\Longrightarrow G_X^{(1)}(1) = \frac{pq}{p^2} = \frac{q}{p}$$

Now, $G_X^{(2)}(t) = pq(-2)(1-tq)^{-3}(-q) = 2pq^2(1-tq)^{-3}$

$$\Longrightarrow G_X^{(2)}(1) = \frac{2pq^2}{p^3} = \frac{2q^2}{p^2}$$

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,

$$G_S(t) = [G_S(t)]^n = \frac{t^n\left(1-t^6\right)^n(1-t)^{-n}}{6^n}$$

$$= \frac{t^n}{6^n}\sum_{j=0}^{n}\binom{n}{j}\left(-t^6\right)^j\sum_{k=0}^{\infty}\binom{n+k-1}{k}t^k$$

$$\Rightarrow G_S(t) = \frac{1}{6^n}\sum_{j=0}^{n}\sum_{k=0}^{\infty}(-1)^j\binom{n}{j}\binom{n+k-1}{k}\,t^{k+6j+n}$$

$$= \sum_{k=0}^{\infty} P\left(S = k + 6j + n\right)\,t^{k+6j+n}$$

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 $k=5$

$$=\frac{1}{6^n}(-1)^0\binom{n}{0}\binom{n+5-1}{5}=\frac{1}{6^n}\binom{n+4}{5}$$

$$\Longrightarrow P(S=n+5)=\frac{1}{6^n}\binom{n+4}{5}$$

P1:

Urns U_1 and U_2 have the following distribution

U_1

x	0	1	2	3
$p_1(x)$	0.4	0.2	0.1	0.3

U_2

y	1	2	3
$p_1(y)$	0.3	0.6	0.1

A ball is drawn from each urn and the numbers X, Y appearing are added. Find the p.m.f of $X + Y$ by using p.g.f.

Solution:

Here X and Y are independent. The p.g.fs of X and Y are given by

$$G_X(t) = \sum_{x=0}^{3} t^x p_1(x) = 0.4 + 0.2t + 0.1t^2 + 0.3t^3$$

and

$$G_Y(t) = \sum_{y=1}^{3} t^y p_2(y) = 0.3t + 0.6t^2 + 0.1t^3$$

respectively. Since X and Y are independent

$$G_{X+Y}(t) = G_X(t)G_Y(t) = (0.4 + 0.2t + 0.1t^2 + 0.3t^3)(0.3t + 0.6t^2 + 0.1t^3)$$

$$= 0.12t + 0.30t^2 + 0.19t^3 + 0.17t^4 + 0.19t^5 + 0.03t^6$$

Thus the p.m.f. of $X + Y$ is given by

r	1	2	3	4	5	6
$P(X + Y = r)$	0.12	0.30	0.19	0.17	0.19	0.03

P2:

Lottery tickets bear numbers from 000000 to 999999. Find the probability that a ticket bears a number whose sum of the first three digits equals the sum of the last three digits.

Solution:

Let the ticket bear the number $Y_1Y_2Y_3Y_4Y_5Y_6$. Obviously, all Ys are identically and independently distributed with

$$p(x) = P\left(Y_j = x\right) = \frac{1}{10}\,, \qquad x = 0, 1, 2, \ldots, 9$$

The p.g.f. of each Y is given by

$$G_Y(t) = \frac{1}{10}(1 + t + \cdots + t^9)$$

$$\Longrightarrow \quad G_Y(t) = \frac{1}{10}\frac{(1-t^{10})}{(1-t)}$$

Let $X_1 = Y_1 + Y_2 + Y_3$ and $X_2 = Y_4 + Y_5 + Y_6$.

Then $G_{X_1}(t) = G_{Y_1}(t)G_{Y_2}(t)G_{Y_3}(t)$

$$= \left(\frac{1}{10}\right)^3 (1 - t^{10})^3(1 - t)^{-3}$$

Similarly $G_{X_2}(t) = \left(\frac{1}{10}\right)^3 (1 - t^{10})^3(1 - t)^{-3}$

Let $G_{X_1}(t) = G_{X_2}(t) = G(t)$

Then $P(X_1 = r) = coef.of\ t^r$ in $G(t)$

and $P(X_2 = r) = coef.of\ t^{-r}$ in $G(t^{-1})$

$\therefore p = P(X_1 = X_2) = coef.of\ t^{\circ}$ in $G(t)G(t^{-1})$

$= coef.of\ t^{\circ}$ in $\left(\frac{1}{10}\right)^3 (1-t^{10})^3(1-t)^{-3}.\left(\frac{1}{10}\right)^3\left(1-\frac{1}{t^{10}}\right)^3\left(1-\frac{1}{t}\right)^{-3}$

$\Longrightarrow p = coef.of\ \ t^{\circ}$ in $\left(\frac{1}{10}\right)^3 (1-t^{10})^6.t^{-27}(1-t)^{-6}$

Now, $(1-t^{10})^6 = 1-\binom{6}{1}t^{10}+\binom{6}{2}t^{20}-\cdots$

and $(1-t)^{-6} = 1+\binom{6}{5}t+\binom{7}{5}t^2+\cdots\binom{5+r}{5}t^r+\cdots$

$\therefore\ \ p = \left(\frac{1}{10}\right)^6\left[\binom{32}{5}-\binom{6}{1}\binom{22}{5}+\binom{6}{2}\binom{12}{5}\right]$

P3:

Find the p.g.f. of $\boldsymbol{P(X > n)}$.

Solution:

The p.g.f. of X is defined by

$G(t) = \sum_{x=0}^{\infty} p(x) t^x$,where $p(x)$ is the p.m.f. of X.

Let $f_n = P(X > n)$ and $p(n) = P(X = n)$

Consider $P(X = n + 1) = P(X > n) - P(X > n + 1)$

$\Longrightarrow \ p(n + 1) = f_n - f_{n+1}$

$\therefore \ t^n f_n - \left(\frac{1}{t}\right) t^{n+1} f_{n+1} = \frac{1}{t} p(n + 1)$

$$\Rightarrow \sum_{n=0}^{\infty} t^n f_n - \frac{1}{t} \sum_{n=0}^{\infty} t^{n+1} f_{n+1} = \frac{1}{t} \sum_{n=0}^{\infty} t^{n+1} p(n+1) \quad \text{... (1)}$$

Let $H(t) = \sum_{n=0}^{\infty} t^n f_n$ ... (2)

Now, $\sum_{n=0}^{\infty} t^{n+1} p(n+1) = tp(1) + t^2 p(2) + t^3 p(3) + \cdots$

$$\Rightarrow \sum_{n=0}^{\infty} t^{n+1} p(n+1) = G(t) - p(0) \quad \text{... (3)}$$

Note that $\sum_{n=0}^{\infty} t^{n+1} f_{n+1} = H(t) - f_0$ ... (4)

But $f_0 = P(X > 0) = 1 - P(X = 0) = 1 - p(0)$ ... (5)

From (4) and (5), we have

$$\sum_{n=0}^{\infty} t^{n+1} f(n+1) = H(t) - 1 + p(0) \qquad \dots (6)$$

On substituting (2), (3) and (6) in (1), we get

$$H(t) - \frac{1}{t}[H(t) - 1 + p(0)] = \frac{1}{t}[G(t) - p(0)]$$

$$tH(t) - H(t) + 1 - p(0) = G(t) - p(0)$$

$$\Longrightarrow\ (t-1)H(t) + 1 = G(t)$$

$$\Longrightarrow H(t) = \frac{G(t)-1}{t-1} \Longrightarrow H(t) = \frac{1-G(t)}{1-t}$$

P4:

Find the p.g.f. of $\boldsymbol{P(X < n)}$.

Solution:

From the relation $P(X < n + 1) - P(X < n) = P(X = n)$,

we get $\emptyset_{n+1} - \emptyset_n = p(n)$ where $\emptyset_n = P(X < n)$

$$\therefore \quad \frac{1}{t}\left(\sum_{n=0}^{\infty} t^{n+1}\phi_{n+1}\right) - \left(\sum_{n=0}^{\infty} t^n \phi_n\right) = \left(\sum_{n=0}^{\infty} t^n p(n)\right)$$

or $t^{-1}[H(t) - \emptyset_0] - H(t) = G(t)$

Since, $\emptyset_0 = P(X < 0) = 0$, the above result is

$$t^{-1}H(t) - H(t) = G(t) \Longrightarrow (1 - t)H(t) = tG(t) \Longrightarrow H(t) = \frac{t.G(t)}{1 - t}$$

3.5. Probability Generating Function

Exercise

1. Let X be a $\mathrm{r.v.}$ with the following p.m.f.

x	1	2	3	4
$p(x)$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{8}$	$\frac{3}{8}$

Find the p.g.f.

2. If $p(x) = P(X = x) = \frac{1}{2^{x-2}}$ for $x = 3, 4, 5, \ldots$. Find the p.g.f. of X

3. Let X be a $r.v.$ with p.g.f. $G_X(t)$. Find the p.g.f. of

   (i) $X + 1$
   (ii) $2X$

4. If $G(t)$ is the p.g.f. of X, then find the p.g.f. of $\frac{X-a}{b}$ where a and b are real constants.

5. Let $G_X(t) = \frac{\left(\frac{1}{3}t+\frac{2}{3}\right)^4}{t}$. What is the range of X? What is its p.m.f.?

Answers

1. $G_X(t) = \frac{1}{3}t + \frac{1}{6}t^2 + \frac{1}{8}t^3 + \frac{3}{8}t^4$

2. $G_X(t) = \frac{t^3}{2-t}$

3. (i) $tG_X(t)$

   (ii) $G_X(t^2)$

4. $t^{-\frac{1}{b}}G\left(t^{\frac{1}{b}}\right)$

5.

x	-1	0	1	2	3
$p(x)$	$\frac{16}{81}$	$\frac{32}{81}$	$\frac{24}{81}$	$\frac{8}{81}$	$\frac{1}{81}$

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) ... (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 $X_1, X_2, \dots, X_n$,

.

$X_{(n)}$ = 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 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 $1 \le i < j \le n$, 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 $X_{(j_1)}, X_{(j_2)}, \dots, X_{(j_k)}$ where $1 \le r_1 < r_2 < \cdots < r_k \le n$ and $1 \le k \le n$ is for $x_1 \le x_2 \le \cdots \le x_k$ given by
$f_{X_{(j_1)},\dots,X_{(j_k)}}(x_1, \dots, x_k)$

$$= \frac{n!}{(j_1 - 1)!\,(j_2 - j_1 - 1)!\dots(j_k - j_{k-1} - 1)!\,(n - j_k)!} \times$$

$$F^{j_1-1}(x_1)[F(x_2) - F(x_1)]^{j_2-j_1-1}\dots[F(x_k) - F(x_{k-1})]^{j_k-j_{k-1}-1} \times$$

$$\left[[1 - F(x_k)]^{n-j_k}\right] f(x_1)f(x_2)\dots f(x_k)$$

6) The j.p.d.f. of $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ is given by

$$f_{X_{(1)},\dots,X_{(n)}}(x_1, \dots, x_n) = \begin{cases} n!\,f(x_1)\dots f(x_n) \;, & -\infty < x_1 < \cdots < x_n < \infty \\ 0 \;, & otherwise \end{cases}$$

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

$f_{X_{(i)},X_{(j)}}(u, v)$

$$= \frac{n!}{(i-1)!(j-i-1)!(n-j)!}[F(u)]^{i-1}[F(v) - F(u)]^{j-i-1}[1 - F(v)]^{n-j}f(u)f(v) \dots .(1)$$

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

$$f_{X_{(1)}}(x) = n[1 - F(x)]^{n-1} f(x) = 4\left(1 - \frac{x}{\theta}\right)^{4-1} \frac{1}{\theta}$$

$$\Rightarrow f_{X_{(1)}}(x) = \frac{4}{\theta}\left(1 - \frac{x}{\theta}\right)^3, 0 < x < \theta$$

The p.d.f of $X_{(3)}$ is given by

$$f_{X_{(3)}}(x) = \frac{4!}{2!\,1!}\left[\frac{x}{\theta}\right]^2 \left(1 - \frac{x}{\theta}\right)^1 \frac{1}{\theta} = \frac{12x^2(\theta - x)}{\theta^4}$$

$$\Rightarrow f_{X_{(3)}}(x) = \frac{12x^2(\theta - x)}{\theta^4}, 0 < x < \theta$$

The p.d.f of $X_{(4)}$ is given by

$$f_{X_{(4)}}(x) = n[F(x)]^{n-1} f(x) = 4\left(\frac{x}{\theta}\right)^3 \frac{1}{\theta} = \frac{4x^3}{\theta^4}$$

$$\Rightarrow f_{X_{(4)}}(x) = \frac{4x^3}{\theta^4}, 0 < x < \theta$$

Example 2: Let $X_1, X_2, \dots, X_n$ be i.i.d.r.v s with common p.d.f .

$$\boldsymbol{f(x)} = \begin{cases} \mathbf{1} & , \quad \mathbf{0} < x < 1 \\ \mathbf{0} & , \ \boldsymbol{otherwise} \end{cases}$$

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

$$f(y_1, y_2) = 3!\,\beta^2\left(1 - e^{-\beta y_2}\right)e^{-\beta y_2}e^{-\beta(y_1+y_2)}\,, 0 < y_1 < \infty\,, 0 < y_2 < \infty \quad \text{.... (1)}$$

The m.p.d.f. of Y_2 is given by

$$f_2(y_2) = 3!\,\beta e^{-2\beta y_2}\left(1 - e^{-\beta y_2}\right), 0 < y_2 < \infty \quad \text{.... (2)}$$

and the m.p.d.f. of Y_1 is given by

$$f_1(y_1) = \beta e^{-\beta y_1},\ 0 < y_1 < \infty \quad \text{.... (3)}$$

From (1), (2) and (3), Y_1 and Y_2 are independent.

Example 4: Let $X_1, X_2, \ldots, X_n$ be a random samples from a population with continuous density. Show that $Y = min(X_1, X_2, \ldots, X_n)$ 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 $F(x) = P(X \leq x) = \int_0^x f(u)du = \lambda \int_0^x e^{-\lambda u} du = 1 - e^{-\lambda x}$

The distribution function of $Y_1 = min(X_1, \ldots, X_n)$ 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 $Y_1 = min(X_1, \ldots, X_n) \sim exp(n\lambda)$

Then $P(Y_1 \leq y) = 1 - e^{-n\lambda y}$

Now, $P(Y_1 \geq y) = 1 - P(Y_1 \leq y) = 1 - \left(1 - e^{-n\lambda y}\right) = e^{-n\lambda y}$
$\Rightarrow P[min(X_1, \ldots, X_n) \geq y] = e^{-n\lambda y}$

$\Rightarrow P[X_1 \geq y\,,\ X_2 \geq y\,, \ldots.., X_n \geq y] = e^{-n\lambda y}$

$\Rightarrow \prod_{i=1}^{n} P(X_i \geq y) = e^{-n\lambda y}$ $\qquad (\because \quad Xs\ are\ i.d.d)$

$\Rightarrow P(X_i \geq y) = e^{-\lambda y} \quad \Rightarrow P(X_i \leq y) = 1 - e^{-\lambda y}$

which is $exp(\lambda)$ distribution. Thus, X_i's are i.d.d $Exp(\lambda)$.

Example 5: For exponential distribution $f(x) = e^{-x}, x \geq 0$, show that the c.d.f. of $X_{(n)}$ in a random sample of size n is $F_n(x) = (1 - e^{-x})^n$. Hence prove that as $n \longrightarrow \infty$, the c.d.f. of $X_n - \ln n$ tends to the limiting form

$$\boldsymbol{exp\big(-exp(-x)\big), -\infty < x < \infty.}$$

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

$$G_n(x) = P\big[X_{(n)} - \ln n \leq x\big] = P\big[X_{(n)} \leq x + \ln n\big]$$

$$= \big[1 - e^{-(x + \ln n)}\big]^n = [1 - e^{-x} e^{-\ln n}]^n$$

$$\implies \quad G_n(x) = \left(1 - \frac{e^{-x}}{n}\right)^n$$

$$\implies \lim_{n\to\infty} G_n(x) = \lim_{n\to\infty} \left(1 - \frac{e^{-x}}{n}\right)^n = e^{-e^{-x}} \qquad \left(\because \lim_{n\to\infty} \left(1 + \frac{x}{n}\right)^n = e^x\right)$$

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 $X_1, X_2, \ldots X_n$ be a random sample, from a population with p.m.f. $p(x_i) = P(X = x_i)$ for $i = 1, 2, \ldots$

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 $X_1, X_2, \dots X_n$ are i.i.d.r.vs with common geometric p.m.f. given by

$$p_k = P(X = k) = pq^{k-1}, \;\; k = 1, 2, \dots, \;\; 0 < p < 1, \; q = 1 - p$$

(i) Find p.m.f. of $X_{(r)}, 1 \le r \le n$ 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 $x \ge 1$ and $r \ge 1$,

$$P\left[X_{(r)} = x\right] = P\left[X_{(r)} \le x\right] - P\left[X_{(r)} \le (x-1)\right]$$

Now $P\left(X_{(r)} \le x\right) = P[at\ least\ \ r\ of\ X\ s\ are\ \le x]$

$$= \sum_{i=1}^{r} \binom{n}{i} \left[P\left(X_1 \le x\right)\right]^i \left[P\left(X_1 > x\right)\right]^{n-i}$$

and $P\left(X_1 \ge x\right) = \sum_{k=x}^{\infty} pq^{k-1} = (1-p)^{x-1} = q^{x-1}$

It follows that, $P\left[X_{(r)} = x\right] = \sum_{i=r}^{n} \binom{n}{i} q^{(x-1)(n-i)} \left[q^{n-i}\left(1-q^x\right)^i - \left(1-q^{x-1}\right)^i\right]$

$x = 1, 2, \dots.$

(ii) Let $n = r = 2$. Then , $P\left[X_{(2)} = x\right] = pq^{x-1}\,(pq^{x-1} + 2 - 2q^{x-1})\,, x \ge 1$

Also, for integers $x, y \ge 1$, we have $P\left[X_{(1)} = x,\; X_{(2)} - X_{(1)} = y\right]$

$$= P\left[X_{(1)} = x, X_{(2)} = x + y\right]$$

$$= P[X_1 = x, X_2 = x + y] + P[X_1 = x + y, X_2 = x]$$

$$= 2pq^{x-1}pq^{x+y-1} = 2pq^{2x-2}pq^{y}$$

$$= P\left(X_{(1)} = x\right)P\left(X_{(2)} = y\right)$$

and $P\left(X_{(1)} = 1,\; X_{(2)} - X_{(1)} = 0\right) = P\left(X_{(1)} = X_{(2)} = 1\right) = p^2$

It follows that $X_{(1)}$ and $X_{(2)} - X_{(1)}$ are independent random variables and, moreover, that $X_{(2)} - X_{(1)}$ a geometric distribution.

AMS570

Order Statistics

1. Definition: Order Statistics of a sample.

Let $X_1, X_2, \ldots, X_n$ be a random sample from a population with p.d.f. f(x). Then,

$X_{(1)} = \min(X_1, X_2, \ldots, X_n)$
$X_{(n)} = \max(X_1, X_2, \ldots, X_n)$
and $X_{(1)} \le X_{(2)} \le \cdots \le X_{(k)} \le \cdots \le X_{(n)}$

2. p.d.f.'s for $X_{(1)}$ and $X_{(n)}$

W.L.O.G.(Without Loss of Generality), let's assume X is continuous.

$$P(X_{(1)} > x) = P(X_1 > x, X_2 > x, \ldots, X_n > x) = \prod_{i=1}^{n} P(X_i > x)$$

$$= \prod_{i=1}^{n} [1 - F_{X_i}(x)]$$

$$F_{X_{(1)}}(x) = 1 - \prod_{i=1}^{n}[1 - F_{X_i}(x)]$$

$$f_{X_{(1)}}(x) = -\frac{d}{dx}\prod_{i=1}^{n}[1 - F_{X_i}(x)] = -\frac{d}{dx}\prod_{i=1}^{n}[1 - F(x)]^n = n[1 - F(x)]^{n-1} f(x)$$

$$F_{X_{(n)}}(x) = P(X_{(n)} \le x) = P(X_1 \le x, X_2 \le x, \ldots, X_n \le x) = \prod_{i=1}^{n} F_{X_i}(x) = [F(x)]^n$$

$$f_{X_{(n)}}(x) = n[F(x)]^{n-1} f(x)$$

Example 1. Let $X_n \overset{i.i.d}{\sim} \exp(\lambda)$, i=1,…,n
Please (1). Derive the MLE of λ
(2). Derive the p.d.f. of $X_{(1)}$
(3). Derive the p.d.f. of $X_{(n)}$

Solutions.

(1).

$$L = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \left(\lambda e^{-\lambda x_i}\right) = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}$$

$$l = \ln L = n\ln\lambda - \lambda \sum_{i=1}^{n} x_i$$

$$\frac{dl}{d\lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0 \Rightarrow \hat{\lambda} = \frac{1}{\overline{\overline{X}}}$$

Is $\hat{\lambda}$ an unbiased estimator of λ? $E\left(\frac{1}{\overline{\overline{X}}}\right) =?$

$$M_{X_i}(t) = \frac{\lambda}{\lambda - t}$$

$$M_{\sum_{i=1}^{n} x_i}(t) = \left(\frac{\lambda}{\lambda - t}\right)^n$$

$$Y = \sum_{i=1}^{n} x_i \sim \text{gamma}(\lambda, n)$$

$$f_{Y-\sum_{i=1}^{n} x_i}(x) = \frac{\lambda}{\Gamma(n)}(\lambda y)^{n-1}e^{-\lambda y}$$

$$\text{Let } Y = \sum_{i=1}^{n} x_i$$

$$E\left(\frac{1}{Y}\right) = \int_0^\infty \frac{1}{y}\frac{\lambda}{(n-1)!}(\lambda y)^{n-1}e^{-\lambda y}dy$$

$$= \frac{\lambda}{n-1}\int_0^\infty \frac{\lambda}{(n-2)!}(\lambda y)^{n-2}e^{-\lambda y}dy$$

$$= \frac{\lambda}{n-1}$$

$$E\left(\frac{1}{\overline{\overline{X}}}\right) = n\left(\frac{\lambda}{n-1}\right) = \frac{n\lambda}{n-1} \neq \lambda$$

$\hat{\lambda}$ is not unbiased

(2). $X_{(1)} = \min(X_1, X_2, \dots, X_n)$

$$P(X_{(1)} > x) = \prod_{i=1}^{n} P(X_i > x) = \prod_{i=1}^{n}[1 - F(x)] = [1 - F(x)]^n$$

$$F_{X_{(1)}}(x) = 1 - [1 - F(x)]^n$$

$$f_{X_{(1)}}(x) = n[1 - F(x)]^{n-1}f(x)$$

$$f(x) = \lambda e^{-\lambda x}, x > 0 \ (\textit{exponential distribution})$$

$$F(x) = \int_0^x f(u)du = \int_0^x \lambda e^{-\lambda u}du = \left[-e^{-\lambda u}\right]\Big|_0^x = 1 - e^{-\lambda x}$$

$$f_{X_{(1)}}(x) = n\lambda e^{-\lambda x}[1 - (1 - e^{-\lambda x})]^{n-1} = n\lambda e^{-\lambda x}(e^{-\lambda x})^{n-1}$$

$$= n\lambda(e^{-\lambda x})^n, x > 0$$

(3). $X_{(n)} = \max(X_1, X_2, \dots, X_n)$

$$F_{X_{(n)}}(x) = P\left(X_{(n)} \le x\right) = \prod_{i=1}^{n} P(X_i \le x) = [F(x)]^n$$

$$f_{X_{(n)}}(x) = n[F(x)]^{n-1}f(x) = n[1 - e^{-\lambda x}]^{n-1}\lambda e^{-\lambda x}, x > 0$$

3. Order statistics are useful in deriving the MLE's.

Example 2. Let X be a random variable with pdf.

$$f(x) = \begin{cases} 1, \text{if } x \in [\theta - \frac{1}{2}, \theta + 1/2] \\ 0, \text{otherwise} \end{cases}$$

Derive the MLE of θ.

Solution.

Uniform Distribution $\Rightarrow$ important!!

$$L = \prod_{i=1}^{n} f(x_i) = \begin{cases} 1, \text{if all } x_i \in [\theta - \frac{1}{2}, \theta + 1/2] \\ 0, \text{otherwise} \end{cases}$$

MLE : max lnL -> max L

means => $\theta - \frac{1}{2} \le X_1 \le \theta + 1/2$

$\theta - \frac{1}{2} \le X_2 \le \theta + 1/2$

…

$\theta - \frac{1}{2} \le X_n \le \theta + 1/2$

Now we re-express the domain in terms of the order statistics as follows:

$$\theta - \frac{1}{2} \le X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)} \le \theta + 1/2$$

$$\theta \le X_{(1)} + \frac{1}{2}$$

$$\theta \ge X_{(n)} - \frac{1}{2}$$

Therefore,

If $\theta \in \left[X_{(n)} - \frac{1}{2}, X_{(1)} + \frac{1}{2}\right]$, then $L = 1$

Therefore, any $\hat{\theta} \in \left[X_{(n)} - \frac{1}{2}, X_{(1)} + \frac{1}{2}\right]$ is an MLE for θ.

4. The pdf of a general order statistic $X_{(j)}$

Let $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$ denote the order statistics of a random sample, $X_1, X_2, \cdots, X_n$, from a continuous population with cdf $F_X(x)$ and pdf $f_X(x)$. Then the pdf of $X_{(j)}$ is

$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!\,(n-j)!} f_X(x)[F_X(x)]^{j-1}[1-F_X(x)]^{n-j}$$

Proof: Let *Y* be a random variable that counts the number of $X_1, X_2, \cdots, X_n$ less than or equal to *x*. Then we have $Y \sim B\big(n, F_X(x)\big)$. Thus:

$$F_{X_{(j)}}(x) = P(Y \geq j) = \sum\nolimits_{k=j}^{n} \binom{n}{k} [F_X(x)]^k [1-F_X(x)]^{n-k}$$

5. The Joint Distribution of Two Order Statistics $X_{(i)}$ *and* $X_{(j)}$

Let $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$ denote the order statistics of a random sample, $X_1, X_2, \cdots, X_n$, from a continuous population with cdf $F_X(x)$ and pdf $f_X(x)$. Then the joint pdf of $X_{(i)}$ and $X_{(j)}$, $1 \leq i < j \leq n$, is

$$f_{X_{(i)},X_{(j)}}(u,v)$$
$$= \frac{n!}{(i-1)!\,(j-1-i)!\,(n-j)!} f_X(u) f_X(v)[F_X(u)]^{i-1}[F_X(v) - F_X(u)]^{j-1-i}[1-F_X(v)]^{n-j}$$
$$for -\infty < u < v < \infty$$

6. Special functions of order statistics

(1) Median (of the sample):

$$\begin{cases} X_{(k+1)}; & if\ n = 2k+1 \\ \dfrac{X_{(k)} + X_{(k+1)}}{2}; & if\ n = 2k \end{cases}$$

(2) Range (of the sample): $X_{(n)} - X_{(1)}$

7. More examples of order statistics

Example 3. **Let X_1,X_2, X_3 be a random sample from a distribution of the continuous type having pdf f(x)=2x, 0<x<1, zero elsewhere.**
(a) compute the probability that the smallest of X_1,X_2, X_3 exceeds the median of the distribution.
(b) If $Y_1 \le Y_2 \le Y_3$ are the order statistics, find the correlation between Y_2 and Y_3.

Answer:
(a)

$$F(x) = P(X_i < x) = x^2;$$

$$\int_0^t 2x dx = \frac{1}{2}; t = \frac{\sqrt{2}}{2}$$

$$P(\min(X_1, X_2, X_3) > t) = P(X_1 > t, X_2 > t, X_3 > t) = P(X_1 > t)P(X_2 > t)P(X_3 > t)$$

$$= [1 - F(t)]^3 = (1 - t^2)^3 = \frac{1}{8}$$

(b)
Please refer to the textbook/notes for the order statistics pdf and joint pdf formula. We have

$$f_{Y_3}(x) = 6 * x^5; 0 < x < 1$$

$$f_{Y_2}(x) = 12 * (x^3 - x^5); 0 < x < 1$$

$$E(Y_3) = 6/7,$$

$$E(Y_2) = 24/35;$$

$$f_{Y_2,Y_3}(y_2, y_3) = 24 * (y_2)^3 * y_3; 0 < \ y_2 \le y_3 < 1$$

$$E(Y_2 Y_3) = \int_0^1 \left[\int_0^{y_3} y_2 * y_3 * 24 * (y_2)^3 * y_3 \, dy_2 \right] dy_3 = \frac{3}{5};$$

$$var(Y_3) = \frac{6}{8} - \left(\frac{6}{7}\right)^2 = \frac{6}{392};$$

$$var(Y_2) = \frac{1}{2} - \left(\frac{24}{35}\right)^2;$$

$$corr(Y_2, Y_3) = \frac{E(Y_2 Y_3) - E(Y_2)E(Y_3)}{\sqrt{var(Y_2) var(Y_3)}} = 0.57$$

Example 4. **Let $Y_1 \le Y_2 \le Y_3$ denote the order statistics of a random sample of size 3 from a distribution with pdf f(x) = 1, 0 < x < 1, zero elsewhere. Let Z = $(Y_1$+$Y_3)/2$ be the midrange of the sample. Find the pdf of Z.**

From the pdf, we can get the cdf : F(x) = x, 0<x<1
Let

$$W = Y_1$$
$$Z = (Y_1 + Y_3)/2$$

The inverse transformation is:

$$Y_1 = W$$
$$Y_3 = 2Z - W$$

The joint pdf of Y_1 and Y_3 is:

$$f(y_1, y_3) = \begin{cases} 6(y_3 - y_1), & 0 < y_1 \leq y_3 < 1 \\ 0, & o.w. \end{cases}$$

We then find the Jacobian: J= -2
Now we can obtain the joint pdf of Z, W:

$$f(z, w) = \begin{cases} |2|6(2z - w - w) = 24(z - w), & 0 < w \leq 2z - w < 1 \\ 0, & o.w. \end{cases}$$

From $0 < w \leq 2z - w < 1$, we have:

$$w > 0;$$
$$w > 2z - 1;$$
$$w \leq z$$

Together they give us the domain of w as:

$$\max(0, 2z - 1) < w \leq z$$

Therefore the pdf of Z (non-zero portion) is:

$$f(z) = \begin{cases} \int_{2z-1}^{z} 24(z - w)\, dw = 12(z - 1)^2, & 2z - 1 > 0 \\ \int_{0}^{z} 24(z - w)\, dw = 12(z^2), & 2z - 1 \leq 0 \end{cases}$$

We also remind ourselves that:

$$0 < z < 1$$

Therefore the entire pdf of the midrange Z is:

$$f(z) = \begin{cases} \displaystyle\int_{2z-1}^{z} 24(z-w)\,dw = 12(z-1)^2, & 1/2 < z < 1 \\ \displaystyle\int_{0}^{z} 24(z-w)\,dw = 12(z^2), & 0 < z \le 1/2 \\ 0, & otherwise \end{cases}$$

Example 5. Let $Y_1 \le Y_2 \le Y_3 \le Y_4$ be the order statistics of a random sample of size n = 4 from a distribution with pdf f(x) = 2x, 0 < x < 1, zero elsewhere.

(a) Find the joint pdf of Y_3 and Y_4.

(b) Find the conditional pdf of Y_3, given $Y_4 = y_4$.

(c) Evaluate $E[Y_3|y_4]$.

Solution:

(a)

$$f_{Y_1,Y_2,Y_3,Y_4,}(y_1,y_2,y_3,y_4) = 4!\,(2y_1)(2y_2)(2y_3)(2y_4)$$

for $0 < y_1 \le y_2 \le y_3 \le y_4 < 1$. We have:

$$f_{Y_3,Y_4}(y_3,y_4) = \int_{y_2=0}^{y_3}\int_{y_1=0}^{y_2} f_{Y_1,Y_2,Y_3,Y_4,}(y_1,y_2,y_3,y_4)\,dy_1dy_2$$
$$= 48{y_3}^5 y_4$$

for $0 < y_3 \le y_4 < 1$

(Note: You can also obtain the joint pdf of these two order statistics by using the general formula directly.)

(b)

$$f_{Y_3|Y_4}(y_3|y_4) = \frac{f_{Y_3,Y_4}(y_3,y_4)}{f_{Y_4}(y_4)} = \frac{48{y_3}^5 y_4}{8{y_4}^7} = \frac{6{y_3}^5}{{y_4}^6}$$

for $0 < y_3 \le y_4$.

(c)

$$E[Y_3|Y_4 = y_4] = \int_0^{y_4} \frac{6{y_3}^6}{{y_4}^6}\,dy_3 = \frac{6y_4}{7}$$

Example 6. Suppose X1, . . . , Xn are iid with pdf f(x; θ) = 2x/θ², 0 < x ≤ θ, zero elsewhere. Note this is a nonregular case. Find:

(a) The mle $\hat{\theta}$ for θ.

(b) The constant c so that E(c*$\hat{\theta}$) = θ.

(c) The mle for the median of the distribution.

Solution:

$$\text{(a)} L(\theta; X) = \prod_{i=1}^{n} \frac{2x_i}{\theta^2} = \frac{2^n \prod_{i=1}^{n} x_i}{\theta^{2n}} \le \frac{2^n \prod_{i=1}^{n} x_i}{[\max(x_i)]^{2n}}$$

$$\text{So } \hat{\theta} = X_{(n)} \qquad X_{(n)} = \max(X_1, \cdots X_n)$$

Dear students: note that this is no typo in the above – the truth is that $\theta \ge X_{(n)}$ – and so the smallest possible value for θ is $X_{(n)}$

$$\text{(b) } F_X(x) = \int_0^x \frac{2t}{\theta^2} dt = \frac{x^2}{\theta^2} \qquad 0 < x \le \theta$$

$$\text{So } F_{X_{(n)}}(x) = \left(\frac{x^2}{\theta^2}\right)^n = \frac{x^{2n}}{\theta^{2n}} \qquad 0 < x \le \theta$$

$$f_{X_{(n)}}(x) = \frac{2nx^{2n-1}}{\theta^{2n}} \qquad 0 < x \le \theta$$

$$E(c\hat{\theta}) = cE(\hat{\theta}) = c\int_0^\theta x \frac{2nx^{2n-1}}{\theta^{2n}} dx = \frac{2nc}{2n+1}\theta = \theta$$

$$\text{So } c = \frac{2n+1}{2n}$$

$$\text{(c) Let} F(x) = \frac{x^2}{\theta^2} = \frac{1}{2}, \text{ then } x = \frac{\theta}{\sqrt{2}}$$

So the median of the distribution is $\frac{\theta}{\sqrt{2}}$

The mle for the median of the distribution is

$$\frac{\widehat{\theta}}{\sqrt{2}} = \frac{X_{(n)}}{\sqrt{2}} = \frac{\sqrt{2}}{2} X_{(n)}$$

Mean Squared Error (M.S.E.)

How to evaluate an estimator?

For unbiased estimators, all we need to do is to compare their variances, the smaller the variance, the better is estimator. Now, what if the estimators are not all unbiased? How do we compare them?

Definition: Mean Squared Error (MSE)
Let T=t(X_1, X_2, …, X_n) be an estimator of $\tau(\theta)$, then the M.S.E. of the estimator T is defined as :

$\mathrm{MSE}_t(\tau(\theta)) = \mathrm{E}\left[(\mathrm{T}-\tau(\theta))^2\right]$: average squared distance from T to $\tau(\theta)$

$=\mathrm{E}\left[(\mathrm{T}-\mathrm{E(T)}+\mathrm{E(T)}-\tau(\theta))^2\right]$

$=\mathrm{E}\left[(\mathrm{T}-\mathrm{E(T)})^2\right]+\mathrm{E}\left[(\mathrm{E(T)}-\tau(\theta))^2\right]+2\mathrm{E}[(\mathrm{T}-\mathrm{E(T)})(\mathrm{E(T)}-\tau(\theta))]$

$=\mathrm{E}\left[(\mathrm{T}-\mathrm{E(T)})^2\right]+\mathrm{E}\left[(\mathrm{E(T)}-\tau(\theta))^2\right]+0$

$=\mathrm{Var(T)}+(\mathrm{E(T)}-\tau(\theta))^2$

Here $|\mathrm{E(T)}-\tau(\theta)|$ is "the bias of T"
If unbiased, $(\mathrm{E(T)}-\tau(\theta))^2=0$.
The estimator has smaller mean-squared error is better.

Example 1. Let $X_1, X_2, \ldots, X_n \overset{\mathrm{i.i.d}}{\sim} N(\mu, \sigma^2)$

M.L.E. for μ is $\hat{\mu}=\bar{X}$; M.L.E. for σ^2 is $\hat{\sigma}^2=\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n}$

1. M.S.E. of $\hat{\sigma}^2$?
2. M.S.E. of S^2 as an estimator of σ^2

Solution.

1.

$\mathrm{MSE}_{\hat{\sigma}^2}(\sigma^2)=\mathrm{E}\left[(\hat{\sigma}^2-\sigma^2)^2\right]=\mathrm{Var}(\sigma^2)+(\mathrm{E}(\hat{\sigma}^2)-\sigma^2)^2$

To get $\mathrm{Var}(\hat{\sigma}^2)$, there are 2 approaches.

a. By the first definition of the Chi-square distribution.

Note $X_i \sim N(\mu,\sigma^2)$; $W=\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\sigma^2}\sim {\chi_{n-1}}^2$, Gamma($\lambda=\frac{1}{2}$, $S=\frac{n-1}{2}$)

$$E(W) = \frac{S}{\lambda} = n-1; \mathrm{Var}(W) = \frac{S}{\lambda^2} = 2(n-1)$$

$$\mathrm{Var}(\widehat{\sigma}^2) = \mathrm{Var}(\tfrac{W}{n}\ \ \sigma^2) = \tfrac{\sigma^4}{n^2}\mathrm{Var}(W) = \tfrac{\sigma^4}{n^2}2(n-1)$$

b. By the second definition of the Chi-squre distribution.

For $Z \sim N(0,1)$, $W=\sum_{i=1}^{n} {Z_i}^2$

$$\mathrm{Var}(Z^2) = E\left[\left(Z^2 - E(Z^2)\right)^2\right]$$

$$= E\left[\left(Z^2 - \left(\mathrm{var}(Z^2) + E(Z)\right)\right)^2\right]$$

$$= E[(Z^2-1)^2]$$

Since $\mathrm{Var}(Z) = E(Z^2) - E(Z) =$ 1 from $Z \sim N(0,1)$, $E(Z^2) = 1 = E[Z^4 - 2E(Z^2) + 1]$

$$= E(Z^4) - 1$$

Calculate the 4th moment of $Z \sim N(0,1)$ using the mgf of Z;

$$M_Z(t) = e^{t^2/2}$$

$$M'_Z(t) = te^{t^2/2}$$

$$M''_Z(t) = te^{t^2/2} + t^2e^{t^2/2}$$

$$M^{(3)}{}_Z(t) = 3te^{t^2/2} + t^2e^{t^2/2}$$

$$M^{(4)}{}_Z(t) = 3e^{t^2/2} + 6t^2e^{t^2/2} + t^4e^{t^2/2}$$

Set $t=0$, $M^{(4)}{}_Z(0) = 3 = E(Z^4)$

$$\mathrm{Var}(Z^2) = 3 - 1 = 2$$

$$\mathrm{Var}(W) = \sum_{i=1}^{n-1} \mathrm{Var}({Z_i}^2) = 2(n-1)$$

$$\widehat{\sigma}^2 = \frac{\sigma^2}{n}W, \mathrm{Var}(\widehat{\sigma}^2) = \frac{\sigma^4}{n^2}2(n-1)$$

$$MSE_{\widehat{\sigma}^2}(\sigma^2) = \mathrm{Var}(\widehat{\sigma}^2) + [E(\widehat{\sigma}^2) - \sigma^2]^2$$

$$= \frac{2(n-1)}{n^2}\sigma^4 + [E\left(\frac{n-1}{n}S^2\right) - \sigma^2]^2$$

$$= \frac{2(n-1)}{n^2}\sigma^4$$

$$+ \left[\frac{n-1}{n}\sigma^2 - \sigma^2\right]^2 \left(\text{we know } E(S^2)\right.$$

$$\left.= \sigma^2\right) = \frac{2n-1}{n^2}\sigma^4$$

The M.S.E. of $\hat{\sigma}^2$ is $\frac{2n-1}{n^2}\sigma^4$

We know S^2 is an unbiased estimator of σ^2

$$E[(S^2-\sigma^2)^2] = Var(S^2) + 0 = Var\left(\frac{\sigma^2 W}{n-1}\right)$$

$$= \left(\frac{\sigma^2}{n-1}\right)^2 var(W) = \frac{2\sigma^4}{n-1}$$

Exercise:

Compare the MSE of $\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n}$ and $\hat{\sigma}^2 = S^2 = \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n-1}$.
Which one is a better estimator (in terms of the MSE)?

Solution.

$$MSE_{\hat{\sigma}^2}(\sigma^2) = E\left[\left(\hat{\sigma}^2-\sigma^2\right)^2\right] = Var\left(\hat{\sigma}^2\right) + \left(E\left(\hat{\sigma}^2\right)-\sigma^2\right)^2$$

Then, we have M.S.E. of $\hat{\sigma}^2$ is $\frac{2n-1}{n^2}\sigma^4$.

$$MSE_{S^2}(\sigma^2) = E[(S^2-\sigma^2)^2] = Var(S^2) + (E(S^2)-\sigma^2)^2$$

$$MSE_{S^2}(\sigma^2) = Var(S^2) + 0 = Var\left(\frac{\sigma^2 W}{n-1}\right) =$$

$$\left(\frac{\sigma^2}{n-1}\right)^2 var(W) = \frac{2\sigma^4}{n-1}$$

$$\because \frac{2n-1}{n^2}\sigma^4 - \frac{2\sigma^4}{n-1} = \frac{1-3n}{n^2(n-1)}\sigma^4 < 0$$

$$\therefore MSE_{\hat{\sigma}^2}(\sigma^2) = \frac{2n-1}{n^2}\sigma^4 < \frac{2\sigma^4}{n-1} = MSE_{S^2}(\sigma^2)$$

$$\therefore \hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n} \text{ is the better estimator}$$

<u>Homework :</u> Read the following chapter/sections from our textbook: *<u>Chapter1, 2, 3 (3.1, 3.2, 3.3)</u>, <u>4 (4.1, 4.2, 4.3, 4.5, 4.6)</u>,5 (5.1, 5.2, 5.3, 5.4), 7(<u>7.1, 7.2.1, 7.2.2, 7.3.1).</u> These are materials covered so far in our class.*

4.1. Order Statistics

Exercise

1. Let $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ be the set of order statistics of independent r.v s with common p.d.f.

$$f(x) = \begin{cases} \beta\, e^{-x\beta} & , \quad x \geq 0 \\ 0 & , \quad otherwise \end{cases}$$

   Show that $X_{(r)}$ and $X_{(s)} - X_{(r)}$ are independent for any $s > r$.

2. Let $X_1, X_2, \dots, X_n$ be i.i.d .r.vswith p.d.f.

$$f(y) = \begin{cases} y^{\alpha} & if \quad 0 < y < 1 \\ 0 & otherwise, \quad \alpha > 0 \end{cases}$$

   Show that $\frac{X_{(i)}}{X_{(n)}}$, $i = 1, 2, \dots, n-1$ and $X_{(n)}$ are independent.

3. Let $X_1, X_2, \dots, X_n$ be i.i.d .r.v s with common p.d.f .

$$f(x) = \alpha \frac{\sigma^{\alpha}}{x^{\alpha+1}}, \ x > \sigma \text{ where } \alpha > 0, \ \sigma > 0$$

   Show that $X_{(1)}$ and $\left(\frac{X_{(2)}}{X_{(1)}}, \dots, \frac{X_{(n)}}{X_{(1)}}\right)$ are independent.

4. Let $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ be the order statistics of n independent r.v.s $X_1, X_2, \dots, X_n$ with common p.d.f.

$$f(x) = \begin{cases} 1 & if \ 0 < x < 1 \\ 0 & otherwise \end{cases}$$

   Show that $Y_1 = \frac{X_{(1)}}{X_{(2)}}, \quad Y_2 = \frac{X_{(2)}}{X_{(3)}}, \dots, Y_{n-1} = \frac{X_{(n-1)}}{X_{(n)}}$ and $Y_n = X_{(n)}$ are independent.

5. An urn contains N identical marbles numbered 1 through N. Form this urn n marbles are drawn, and let $X_{(n)}$ be the largest number drawn. Show that

$$P\left[X_{(n)} = k\right] = \frac{\binom{k-1}{n-1}}{\binom{N}{n}}, \quad k = n, n+1, \dots, N$$

and $E\left(X_{(n)}\right) = \frac{n(N+1)}{(n+1)}$.

P1:

Find the j.p.d.f of extremes ($i.e.$, $X_{(1)}$ and $X_{(n)}$).

P1:

Find the j.p.d.f of extremes ($\boldsymbol{i.e.}$, $\boldsymbol{X_{(1)}}$ and $\boldsymbol{X_{(n)}}$).

Solution:

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 $1 \leq i < j \leq n$ and $-\infty < u < v < \infty$.

Let $i = 1$ and $j = n$. Then the j.p.d.f. of $X_{(1)}$ and $X_{(n)}$ is given by

$$f_{X_{(1)},X_{(n)}}(u,v) = n(n-1)[F(v)-F(u)]^{n-2}f(u)f(v), \quad -\infty < u < v < \infty$$

P2:

Find the j.p.d.f. of the mid range $M = \frac{1}{2}\left[X_{(1)} + X_{(2)}\right]$.

P2:

Find the j.p.d.f. of the mid range $\boldsymbol{M = \frac{1}{2}[X_{(1)} + X_{(2)}]}$.

Solution:

Let $x = u$ and $y = \frac{1}{2}(u + v)$.Then $u = x$ and $v = 2y - x$ and the jacobian of transformation is given by

$$J = \frac{\partial(u,v)}{\partial(x,y)} = \begin{vmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ -1 & 2 \end{vmatrix} = 2 \text{ and } |J| = 2$$

$X_{(1)}$ and $X_{(2)}$ given in P_1, the j.p.d.f. of $X_{(1)}$ and M is given by

$$f_{X_{(1)},M}(x,y) = f_{X_{(1)},X_{(n)}}(x,y)|J|$$

$$= 2n(n-1)[F(2y-x) - F(x)]^{n-2} f(x) f(2y-x)$$

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

$$f_M(y) = \int_{-\infty}^{\infty} f_{X_{(1)},M}(x,y)\,dx$$

$$= 2n(n-1)\int_{-\infty}^{\infty} \left[F(2y-x) - F(x)\right]^{n-2} f(x) f(2y-x)\,dx$$

P3:

Find the conditional probability density function of $X_{(i)}$ given $X_{(j)}$ for $1 \leq i < j \leq n$.

P3:

Find the conditional probability density function of $\boldsymbol{X_{(i)}}$ given $\boldsymbol{X_{(j)}}$ for $\mathbf{1} \leq \boldsymbol{i} < j \leq n$.

Solution:

The c.p.d.f. of $X_{(i)}$ given $X_{(j)}$ is given by

$$f_{X_{(i)}|X_{(j)}}(x|y) = \frac{f_{X_{(i)},X_{(j)}}(x,y)}{f_{X_{(j)}}(y)}$$

$$= \frac{\frac{n!}{(i-1)!(j-i-1)!(n-j)!}[F(x)]^{i-1}[F(y)-F(x)]^{j-i-1}[1-F(y)]^{n-j}\,f(x)f(y)}{\frac{n!}{(j-1)!(n-j)!}[F(y)]^{j-1}[1-F(y)]^{n-j}f(y)}$$

$$= \frac{(j-1)!}{(i-1)!(j-i-1)!}[F(x)]^{i-1}[F(y)-F(x)]^{j-i-1}[F(y)]^{1-j}f(x) \text{ for } y \geq x$$

P4:

Find the conditional p.d.f. of $X_{(i)}$ given $X_{(i+1)}$

P4:

Find the conditional p.d.f. of $\boldsymbol{X_{(i)}}$ given $\boldsymbol{X_{(i+1)}}$

Solution:

Let $j = i + 1$ in P3. Then

$f_{X_{(i)}|X_{(j+1)}}(x|y) = i[F(x)]^{i-1}[F(y)]^{-i}f(x)$

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 $\{X_1, X_2, \dots\}$ 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 $\boldsymbol{X_1, X_2, \dots, X_n}$ 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 $\boldsymbol{X_{(n)} = max(X_1, \dots, X_n)}$. 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 $\boldsymbol{X_n} \xrightarrow{\boldsymbol{L}} \boldsymbol{X}$ 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

$$F(x) = \begin{cases} 0 & , \quad x < 2 \\ 1 & , \quad x \geq 2 \end{cases}$$

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

$$P(X_n = 0) = 1 - \frac{1}{n} \text{ and } P(X_n = 1) = \frac{1}{n}\ ,\quad n = 1, 2, \dots$$

Show that $X_n \xrightarrow{q.m} 0$ but $X_n \overset{a.s}{\not\longrightarrow} 0$

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, $P(X_n = 0\ \ for\ every\ m \le n \le n_0) = \prod_{n=m}^{n_0}\left(1 - \frac{1}{n}\right) = \frac{m-1}{n_0}$ which converges to zero as $n \longrightarrow \infty$ for all values of m. Thus, $X_n \overset{a.s}{\not\longrightarrow} 0$

Example 7: Let $\{X_n\}$ be a sequence of independent r.vs defined by

$$P(X_n = 0) = 1 - \frac{1}{n^r} \text{ and } P(X_n = n) = \frac{1}{n^r}\ , r \ge 2,\ n = 1, 2, \dots$$

Show that $X_n \xrightarrow{a.s} 0$ but $X_n \overset{r}{\not\longrightarrow} 0$.

Solution: We have $P(X_n = 0\ \ for\ \ m \le n \le n_0) = \prod_{n=m}^{n_0}\left(1 - \frac{1}{n^r}\right)$

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, $X_n \overset{r}{\not\longrightarrow} 0$

Thus, $X_n \xrightarrow{a.s} 0$ but $X_n \overset{r}{\not\longrightarrow} 0$

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

$$\bigcup_{j=n}^{\infty}\left\{\left|X_j\right| > \varepsilon\right\} = \left\{\left|X_n\right| > \varepsilon\right\}$$

Choosing $n > \frac{1}{\epsilon}$, we see that

$$P\left[\bigcup_{j=n}^{\infty}\left\{\left|X_j\right| > \varepsilon\right\}\right] = P\left(\left\{\left|X_n\right| > \varepsilon\right\}\right) \leq P\left(|X_n| > \frac{1}{n}\right) = 0 \text{ as } n \longrightarrow \infty$$

$$\Rightarrow \lim_{n\to\infty} P\left[\bigcup_{j=n}^{\infty}\left\{\left|X_j\right| > \varepsilon\right\}\right] = 0 \Longrightarrow X_n \xrightarrow{a.s} 0$$

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) $X_n + Y_n \xrightarrow{L} X + c$

(ii) $X_n Y_n \xrightarrow{L} cX$

(iii) $\frac{X_n}{Y_n} \xrightarrow{L} \frac{X}{C}$ **if** $c \neq 0$

An example presented in **P4** as an application of **Slutsky's theorem.**

4.2. Convergence of sequences of random variables

Exercise:

1. Let $X_1, X_2, ..$ be a sequence of r.vs with corresponding d.fs given by

$$F_n(x) = \begin{cases} 0 & , & x < n \\ \frac{x+n}{2n} & , & -n \leq x < n \\ 0 & , & x \geq n \end{cases}$$

   Does F_n converge to some d.f.

2. Let$X_1, X_2, ..$ be a i.i.d $U(0, \theta)$ r.vs. Let $X_{(1)} = min(X_1, X_2, \ldots, \mathrm{X_n})$ and consider the sequence $Y_n = nX_{(1)}$. Does Y_n converge in distribution to some r.v. Y? If so, find the d.f. of r.v. Y.

3. Let $X_1, X_2, ..$ be i.i.d. r. vs with continuous d.f. F .Let $X_{(n)} = max(X_1, X_2, .. \mathrm{X_n})$ and consider the sequence of r.vs $Y_n = n\left[1 - F\left(X_{(n)}\right)\right]$.Find the limiting d.f. of of Y_n.

4. Let $X_1, X_2, ..$ be a sequence of i.i.d r.vs with common p.d.f

$$f(x, \theta) = \begin{cases} e^{-x+\theta}, & if \ \ x \geq \theta \\ 0 \quad , & if \ \ x < \theta \end{cases}$$

   write $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

   (a) Show that $\bar{X}_n \xrightarrow{P} 1 + \theta$

   (b) Show that $\min\{X_1, X_2, .. \mathrm{X_n}\} \xrightarrow{P} \theta$

5. Let $X_1, X_2, ..$ be i.i.d $U[0, \theta]$ r.vs. Show that $\max\{X_1, X_2, .., \mathrm{X_n}\} \xrightarrow{P} \theta$

Answers:

1. No
2. Yes. $Y_n \longrightarrow Y$ where $F(y) = \begin{cases} 0 & if \quad y < 0 \\ 1 - e^{-y/\theta} & if \quad y \geq 0 \end{cases}$
3. $F(y) = \begin{cases} 0 & if \quad y \leq 0 \\ 1 - e^{-y} & if \quad y > 0 \end{cases}$

P1:

The j. p. m. f of X and X_n is given by

X \ X_n	0	0	$P(X=x)$
0	0	$\frac{1}{2}$	$\frac{1}{2}$
1	$\frac{1}{2}$	0	$\frac{1}{2}$
$P(X_n = x)$	$\frac{1}{2}$	$\frac{1}{2}$	1

Show that $X_n \xrightarrow{d} X$ but $X_n \overset{P}{\not\longrightarrow} 0$

P1:

The $\mathbf{j.p.m.f}$ of $\boldsymbol{X}$ and $\boldsymbol{X_n}$ is given by

X \ X_n	$\mathbf{0}$	$\mathbf{0}$	$\boldsymbol{P(X=x)}$
$\mathbf{0}$	$\mathbf{0}$	$\frac{\mathbf{1}}{\mathbf{2}}$	$\frac{\mathbf{1}}{\mathbf{2}}$
$\mathbf{1}$	$\frac{\mathbf{1}}{\mathbf{2}}$	$\mathbf{0}$	$\frac{\mathbf{1}}{\mathbf{2}}$
$\boldsymbol{P(X_n=x)}$	$\frac{\mathbf{1}}{\mathbf{2}}$	$\frac{\mathbf{1}}{\mathbf{2}}$	$\mathbf{1}$

Show that $\boldsymbol{X_n \xrightarrow{d} X}$ but $\boldsymbol{X_n \overset{P}{\not\longrightarrow} 0}$

Solution:

The p.m.f of X_n is $P(X_n=0)=\frac{1}{2}=P(X_n=1)$ and its c.d.f is given by

$$F_n(x)=\begin{cases}0 & , \quad x<0\\ \frac{1}{2} & , \ 0\le x<1\\ 1 & , \quad x\ge 1\end{cases}$$

The p.m.f of X is $P(X=0)=\frac{1}{2}=P(X=1)$ and its c.d.f is given by

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

Thus, $F_n(x)\xrightarrow{w}F(x)$ as $n\longrightarrow\infty$. That is $X_n\xrightarrow{d}X$

But $P\left[|X_n-X|>\frac{1}{2}\right]=P[|X_n-X|=1]$

$$=P[X_n=0,X=1]+P[X_n=1,X=0]$$

$$=\frac{1}{2}+\frac{1}{2}=1\not\longrightarrow 0 \text{ as } n\longrightarrow\infty$$

and hence $X_n\overset{P}{\not\longrightarrow}0$. Therefore, $X_n\xrightarrow{d}X$ but $X_n\overset{P}{\not\longrightarrow}0$

P2:

Let $\{X_n\}$ **be a sequence of r.vs defined by** $\boldsymbol{P(X_n = 0) = 1 - \frac{1}{n^r}}$ **and** $\boldsymbol{P(X_n = n) = \frac{1}{n^r}}$ $\boldsymbol{, r > 0, n = 1, 2, \dots}$ **.**

Show that $(\boldsymbol{i})$ $\boldsymbol{X_n \xrightarrow{P} 0}$ **and** $(\boldsymbol{ii})$ $\boldsymbol{X_n \overset{r}{\not\rightarrow} 0}$

P2:

Let $\{X_n\}$ be a sequence of r.vs defined by $P(X_n = 0) = 1 - \frac{1}{n^r}$ and $P(X_n = n) = \frac{1}{n^r}$, $r > 0, n = 1, 2, \dots$.

Show that (i) $X_n \xrightarrow{P} 0$ and (ii) $X_n \not\xrightarrow{r} 0$

Solution:

(i) We have

$$P(|X_n - 0| > \epsilon) = P(X_n = n) = \frac{1}{n^r}$$

Therefore $P(|X_n - 0| > \epsilon) \longrightarrow 0$ as $n \longrightarrow \infty$. Thus, $X_n \xrightarrow{P} 0$

(ii) We have $E(|X_n - 0|^r) = E(|X_n|^r) = 0^r \left(1 - \frac{1}{n^r}\right) + n^r \left(\frac{1}{n^r}\right) = 1$

$$\Rightarrow E(|X_n - 0|^r) = 1$$
$$\Rightarrow \lim_{n\to\infty} E(|X_n - 0|^r) = 1$$
$$\Rightarrow X_n \not\xrightarrow{r} 0$$

Thus $X_n \xrightarrow{P} 0$ but $X_n \not\xrightarrow{r} 0$

P3:

For each positive integer n there exist integers m and k (uniquely determined) such that

$$n = 2^k + m\,, 0 \le m < 2^k\,, k = 0, 1, 2, \ldots$$

Thus, for $n = 1$, we have $k = 0$ and $m = 0$; for $n = 5$, we have $k = 2$ and $m = 1$; and so on. Define r.vs X_n for $n = 1, 2, \ldots n$ on $\Omega = [0, 1]$ by

$$X_n(w) = \begin{cases} 2^k\,, & \frac{m}{2^k} \le w < \frac{m+1}{2^k} \\ 0\,, & \text{otherwise} \end{cases}$$

Show that $X_n \xrightarrow{P} 0$ but $X_n \overset{a.s}{\not\longrightarrow} 0$.

P3:

For each positive integer n there exist integers m and k (uniquely determined) such that

$$\boldsymbol{n = 2^k + m\,, 0 \le m < 2^k\,, k = 0, 1, 2, \dots}$$

Thus, for $n = 1$, we have $k = 0$ and $m = 0$; for $n = 5$, we have $k = 2$ and $m = 1$; and so on. Define r.vs X_n for $n = 1, 2, \dots n$ on $\Omega = [0, 1]$ by

$$\boldsymbol{X_n(w) = \begin{cases} 2^k\,, & \frac{m}{2^k} \le w < \frac{m+1}{2^k} \\ 0\,, & \text{otherwise} \end{cases}}$$

Show that $X_n \xrightarrow{P} 0$ but $X_n \overset{a.s}{\not\longrightarrow} 0$.

Solution:

Let the probability distribution of X_n be given by

$$P(I) = \text{length of the interval } I \subseteq \Omega.$$

Thus, $P(X_n = 2^k) = \frac{1}{2^k}$ and $P(X_n = 0) = 1 - \frac{1}{2^k}$

The limit $\lim_{n\to\infty} X_n(w) = \lim_{k\to\infty} X_n(w)$ does not exist for any $w \in \Omega$, so that X_n does not converge almost surely. But

$$P(|X_n - 0| > \epsilon) = P(|X_n| > \epsilon) = P(X_n > \epsilon)$$

$$= \begin{cases} 0\,, & \epsilon \ge 2^k \\ \frac{1}{2^k}\,, & 0 < \epsilon < 2^k \end{cases}$$

and we see that

$$P(|X_n - 0| > \epsilon) \longrightarrow 0 \text{ as } n \text{ (and hence } k) \longrightarrow \infty.$$

Thus, $X_n \xrightarrow{P} 0$ but $X_n \overset{a.s}{\not\longrightarrow} 0$.

P4:

Let $X_1, X_2, \ldots$ be i.i.d.r.vs with common p.d.f. $N(0,1)$. Determine the limiting distribution of the r.v.

$$W_n = \sqrt{n}\left(\frac{X_1 + \cdots + X_n}{X_1^2 + \cdots + X_n^2}\right)$$

P4:

Let $X_1, X_2, \ldots$ be i.i.d.r.vs with common p.d.f. $N(0,1)$. Determine the limiting distribution of the r.v.

$$\boldsymbol{W_n = \sqrt{n}\left(\frac{X_1+\cdots+X_n}{X_1^2+\cdots+X_n^2}\right)}$$

Solution:

Let $U_n = \frac{1}{\sqrt{n}}(X_1+\cdots+X_n)$ and $V_n = \frac{X_1^2+X_2^2+\cdots+X_n^2}{n}$. Then $W_n = \frac{U_n}{V_n}$

Since each $X \sim N(0,1)$, the m.g.f. of X is given by $M_X(t) = e^{\frac{t^2}{2}}$

Then the m.g.f. of U_n is given by

$$M_{U_n}(t) = E\left[e^{tU_n}\right] = E\left[e^{\left(\frac{t}{\sqrt{n}}\right)\sum_{i=1}^{n} X_i}\right]$$

$$= \prod_{i=1}^{n} M_{X_i}\left(\frac{t}{\sqrt{n}}\right) \qquad (\because X\ s \text{ are independent})$$

$$= \prod_{i=1}^{n} e^{t^2/2n} \qquad (\because X\ s \text{ are identically distributed})$$

$$= e^{\frac{t^2}{2n}}.e^{\frac{t^2}{2n}}.e^{\frac{t^2}{2n}}.e^{\frac{t^2}{2n}}\ldots.e^{\frac{t^2}{2n}}\ (n\ times) = e^{\frac{t^2}{2}}$$

$\Longrightarrow M_{U_n}(t) = e^{\frac{t^2}{2}}$ which is the m.g.f. of $N(0,1)$ r.v. By uniqueness of m.g.f. , $U_n \sim N(0,1)$ $i.e.,$ $U_n \xrightarrow{L} Z$, where $Z \sim N(0,1)$.

Next, we find the m.g.f. of V_n and identify its probability distribution.

First, the m.g.f. of X^2 is given by

$$M_{X^2}(t)=E\left[e^{tx^2}\right]=\int_{-\infty}^{\infty}e^{tx^2}f(x)dx=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{tx^2}e^{-\frac{1}{2}x^2}dx=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{1}{2}x^2(1-2t)}dx$$

Let $\left(\sqrt{1-2t}\right)x=y \Longrightarrow dx=(1-2t)^{-\frac{1}{2}}dy$. Then

$$M_{X^2}(t)=\left\{\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-y^2}dy\right\}(1-2t)^{-1/2}=(1-2t)^{-1/2}$$

($\because$ total probability of $N(0,1)$ is 1)

$\Longrightarrow M_{X^2}(t)=(1-2t)^{-\frac{1}{2}}$

Next, $M_{V_n}(t)=E[e^{tV_n}]=E\left[e^{\left(\frac{t}{n}\right)(X_1^2+\cdots+X_n^2)}\right]$

$$=\prod_{i=1}^{n}E\left[e^{\left(\frac{t}{n}\right)X_i^2}\right]$$ ($\because$ X s are independent)

$$=\prod_{i=1}^{n}M_{X^2}\left(\frac{t}{n}\right)$$

$$=\prod_{i=1}^{n}\left(1-\frac{2t}{n}\right)^{-\frac{1}{2}}=\left(1-\frac{2t}{n}\right)^{-\frac{n}{2}}$$

$\Longrightarrow$ $M_{V_n}(t)=\left(1-\frac{2t}{n}\right)^{-\frac{n}{2}}$, $t<\frac{n}{2}$,which is the m.g.f. of a Gamma distribution with two parameters $\alpha=\frac{n}{2}$ and $\beta=\frac{2}{n}$

The $p.d.f.$ of Gamma variate with two parameters (α,β) is defined by

$$f(x)=\frac{1}{\sqrt{\alpha}}\frac{1}{\beta^{\alpha}}e^{-\frac{1}{\beta}x}x^{\alpha-1} \text{ for } x>0\,,\ \alpha>0,\beta>0$$

Notation: $X\sim G(\alpha,\beta)$

If $X\sim G(\alpha,\beta)$ then its m.g.f. is given by $M_X(t)=(1-pt)^{-\alpha}$

Mean= $E(X)=\alpha\beta$ and variance $=V(X)=\beta^2\alpha$

The variance of V_n is given by $V(V_n) = \beta^2\alpha = \left(\frac{2}{n}\right)^2 \frac{n}{2} = \frac{2}{n}$

We have for any $\epsilon > 0$,

$$P\{|V_n - 1| > \epsilon\} \leq \frac{V(V_n)}{\epsilon^2} \qquad \text{(By chebychev's inequality)}$$

$$= \frac{2}{n\epsilon^2} \longrightarrow 0 \text{ as } n \longrightarrow \infty$$

Thus, $V_n \xrightarrow{\boldsymbol{P}} 1$. We have thus shown that $U_n \xrightarrow{\boldsymbol{L}} Z$ and $V_n \xrightarrow{\boldsymbol{P}} 1$

If follows by Slutsky's theorem (iii)

$$W_n = \frac{U_n}{V_n} \xrightarrow{\boldsymbol{L}} \frac{Z}{1} = Z, \text{ where } Z \text{ is } N(0,1)$$

Hence, $W_n \sim N(0,1)$.

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

$$\lim_{n\longrightarrow\infty} P[\overline{X_n} \le k] = \begin{cases} 0 & , if \;\; k < \mu \\ 1 & , if \;\; k > \mu \end{cases}$$

Proof: Since WLLN holds for $\{X_n\}$, we have

$$\lim_{n\longrightarrow\infty} P[|\overline{X_n} - \mu| < \epsilon] = 1 \Longrightarrow \lim_{n\longrightarrow\infty} P[|\overline{X_n} - \mu| \ge \epsilon] = 0 \qquad \dots (1)$$

Since $\{\overline{X_n} \le \mu - \epsilon\} \subset \{|\overline{X_n} - \mu| \ge \epsilon\}$, we have

$$P(\overline{X_n} \le \mu - \epsilon) \le P(|\overline{X_n} - \mu| \ge \epsilon)$$

$$\Longrightarrow \lim_{n\longrightarrow\infty} P(\overline{X_n} \le \mu - \epsilon) \le \lim_{n\longrightarrow\infty} P(|\overline{X_n} - \mu| \ge \epsilon)$$

$$\Longrightarrow \lim_{n\longrightarrow\infty} P(\overline{X_n} \le \mu - \epsilon) = 0$$

$\Longrightarrow \lim_{n\longrightarrow\infty} P(\overline{X_n} \le k) = 0$, where $k = \mu - \epsilon$, $i.e., k < \mu$ since $\epsilon > 0$

$$\Longrightarrow \lim_{n\longrightarrow\infty} P(\overline{X_n} \le k) = 0 \;\; if \; k < \mu$$

Further, $P(\overline{X_n} \le \mu + \epsilon) + P(|\overline{X_n} - \mu| > \epsilon) \ge 1$, since the region is larger than sample space covered.

$$\Longrightarrow \lim_{n\longrightarrow\infty} P(\overline{X_n} \le \mu + \epsilon) \ge 1 \;\; \left(\because \lim_{n\longrightarrow\infty} P(|\overline{X_n} - \mu| > \epsilon) = 0\right)$$

$$\Longrightarrow \lim_{n\longrightarrow\infty} P(\overline{X_n} \le \mu + \epsilon) = 1$$

$\Longrightarrow \lim_{n\longrightarrow\infty} P(\overline{X_n} \le k) = 1$ where $k = \mu + \epsilon \; i.e., k > \mu$ since $\epsilon > 0$

$\Longrightarrow \lim_{n\longrightarrow\infty} P(\overline{X_n} \le k) = 1$ if $k > \mu$

Thus, $\lim_{n\longrightarrow\infty} P(\overline{X_n} \le k) = \begin{cases} 0\,, & k < \mu \\ 1\,, & k > \mu \end{cases}$

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 $E(X_i) = \mu < \infty, i = 1,2, \ldots$, 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

$$\frac{{X_1}^2 + {X_2}^2 + \cdots + {X_n}^2}{n} \xrightarrow{P} c$$

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 $S_n = {X_1}^2 + {X_2}^2 + \cdots + {X_n}^2$. Then

$E(S_n) = nE({X_1}^2)$ $\qquad$ ($\because$ Xs are i.i.d.r.vs)

$$= n\left[V(X_1) + \left(E(X_1)\right)^2\right]$$

$$\Longrightarrow E(S_n) = n(\sigma^2 + \mu^2)$$

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 $\boldsymbol{X_k (k = 1, 2, \dots)}$ assume the value $\boldsymbol{2^{r-2\ln r}}$ with probability $\boldsymbol{\frac{1}{2^r}}$, examine if the WLLN holds for the sequence $\boldsymbol{\{X_k\}}$.

Solution:

$$E\left(X_k\right) = \sum_{r=1}^{\infty} 2^{r-2\ln r} \cdot \frac{1}{2^r} = \sum_{r=1}^{\infty} \left(2^{-2}\right)^{\ln r} = \sum_{r=1}^{\infty} \left(\frac{1}{4}\right)^{\ln r}$$

$$= \sum_{r=1}^{\infty} (r)^{\ln\left(\frac{1}{4}\right)} \left(\because a^{\ln n} = n^{\ln a}\right)$$

$$= \sum_{r=1}^{\infty} \left(\frac{1}{r}\right)^{\ln 4} \textit{ converges since } \ln 4 = 1.39 > 1 \quad \left(\sum_{n=1}^{\infty} \frac{1}{n^p} \textit{ converges if } p > 1\right)$$

$\textit{Thus } E\left(X_k\right) < \infty$

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 $\boldsymbol{\{X_n\}}$ be a sequence of i.i.d $\boldsymbol{U(0, 1)}$ r.vs. For the geometric mean $\boldsymbol{G_n = (X_1 . X_2 . \dots . X_n)^{\frac{1}{n}}}$, show that $\boldsymbol{G_n \xrightarrow{P} c}$ where $\boldsymbol{c}$ is some constant. Find $\boldsymbol{c}$.

Solution: Let $Y = -\ln X$ where $X \sim U(0,1)$. The c.d.f. of Y is given by

$$F_Y(y) = P(Y \leq y) = P(-\ln X \leq y) = P(X \geq e^{-y}) = \int_{e^{-y}}^{1} 1\, dx = 1 - e^{-y}$$

$\Longrightarrow F_Y(y) = 1 - e^{-y}$ 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

$$\sum_{i=1}^{n} \frac{Y_i}{n} \xrightarrow{P} E(Y_1) = 1 \qquad \dots(1)$$

$$But \ \ln G_n = \sum_{i=1}^{n} \ln \frac{X_i}{n} = -\sum_{i=1}^{n} \frac{Y_i}{n}$$

$$\Longrightarrow \sum_{i=1}^{n} \frac{Y_i}{n} = -\,ln\, G_n \qquad \dots(2)$$

From (1) and (2), we have

$$-\,ln\, G_n \xrightarrow{P} 1 \quad i.e., \quad G_n \xrightarrow{P} e^{-1}$$

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 $S_n = \sum_{k=1}^{n} X_k$. 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 $\alpha < \frac{1}{2}$.

4.3. Weak Law of Large Numbers

Exercise:

1) For the following sequence of independent r.vs, does the $WLLN$ hold?
   a) $P(X_k = \pm k) = \frac{1}{2\sqrt{k}}$, $P(X_k = 0) = 1 - \frac{1}{\sqrt{k}}$
   b) $P(X_k = \pm 2^k) = \frac{1}{2^{2k+1}}$, $P(X_k = 0) = 1 - \frac{1}{2^{2k}}$
   c) $P\left(X_k = \pm \frac{1}{k}\right) = \frac{1}{2}$
   d) $P\left(X_k = \pm \sqrt{k}\right) = \frac{1}{2}$

2) Examine if $WLLN$ holds for the sequence $\{X_i\}$ of i.i.d.r.vs with

$$P[x_i = (-1)^{k-1}.k] = \frac{6}{\pi^2 k^2}, \quad k = 1, 2, 3, \dots, \ \ i = 1, 2, \dots.$$

   Hint: $\log_e 2 = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \cdots$

3) Let $X_1, X_2, \dots, X_n$ be jointly normal with $E(X_i) = 0, E(X_i)^2 = 1$ for all i and

$$cov(X_i, X_j) = \begin{cases} \rho \ , & |j-i| = 1 \\ 0 \ , & \text{otherwise} \end{cases}$$

   Examine if $WLLN$ holds for the sequence $\{X_n\}$.

ANSWERS

1)
   a) No
   b) Yes
   c) Yes
   d) No
   e) No

2) Yes

3) Yes

P1:

A variate X_k has the distribution

$$P(X_k = 0) = 1 - \left(\frac{2}{3^{2k+2}}\right), \; P\left(X_k = \pm 3^k\right) = 3^{-(2k+2)}$$

If $\{X_k\}$ is a sequence of independent r.vs, then show that $\{X_k\}$ obeys $WLLN$

P1:

A variate X_k has the distribution

$$P(X_k = 0) = 1 - \left(\frac{2}{3^{2k+2}}\right),\ P\left(X_k = \pm 3^k\right) = 3^{-(2k+2)}$$

If $\{X_k\}$ is a sequence of independent r.vs, then show that $\{X_k\}$ obeys $WLLN$

Solution:

Here $E(X_k) = 3^k . 3^{-(2k+2)} - 3^k . 3^{-(2k+2)} = 0$ and

$$V(X_k) = E(X_k^2) = 3^{2k} 3^{-(2k+2)} + 3^{2k}\ 3^{-(2k+2)}$$

$$= 2. 3^{2k} 3^{-(2k+1)} = \frac{2}{9}$$

Let $S_n = \sum_{k=1}^{n} X_k$. Then

$$B_n = V(S_n) = \sum_{k=1}^{n} V(X_k) \qquad (\because\ X_k\ s \text{ are independent})$$

$$= \sum_{k=1}^{n} \frac{2}{9} = \frac{2n}{9}$$

and hence $\frac{B_n}{n^2} = \frac{2n}{9n^2} = \frac{2}{9n} \longrightarrow 0$ as $n \longrightarrow 0$

Since $\frac{B_n}{n^2} \longrightarrow 0$ as $n \longrightarrow \infty$, $\{X_k\}$ obeys the $WLLN$.

P2:

Let $\{X_n\}$ be a sequence of independent r.vs such that

$$P(X_n = \pm 1) = \frac{1}{2}(1 - 2^{-n}), \;\; P(X_n = \pm 2^{-n}) = 2^{-n-1}$$

Does the $WLLN$ hold for this sequence?

P2:

Let $\{X_n\}$ be a sequence of independent r.vs such that

$$P(X_n = \pm 1) = \frac{1}{2}(1 - 2^{-n}), \;\; P(X_n = \pm 2^{-n}) = 2^{-n-1}$$

Does the $WLLN$ hold for this sequence?

Solution:

Here $E(X_n) = \frac{1}{2}(1 - 2^{-n}) - \frac{1}{2}(1 - 2^{-n}) + 2^{-n}.2^{-n-1} - 2^{-n}.2^{-n-1} = 0$

and $V(X_n) = E(X_n^2) = \frac{1}{2}(1 - 2^{-n}) + \frac{1}{2}(1 - 2^{-n}) + 2^{-2n}.2^{-n-1} + 2^{-2n}.2^{-n-1}$

$$= (1 - 2^{-n}) + 2^{-3n}$$

$\Longrightarrow V(X_n) = 2^{-3n} - 2^{-n} + 1$

Let $S_n = \sum_{i=1}^{n} X_i$. Then

$V(S_n) = \sum_{i=1}^{n} V(X_i)$ (∵ $X_i\, s$ are independent)

$$= \sum_{i=1}^{n}\left(2^{-3i} - 2^{-i} + 1\right) = \sum_{i=1}^{n} 2^{-3i} - \sum_{i=1}^{n} 2^{-i} + n$$

$\Longrightarrow \; B_n = V(S_n) = \frac{1}{7}(1 - 8^{-n}) - (1 - 2^{-n}) + n = -\frac{1}{78^n} + \frac{1}{2^n} - \frac{6}{7} + n$

$\Longrightarrow \frac{B_n}{n^2} = -\frac{1}{7.n^2.8^n} + \frac{1}{n^2.2^n} - \frac{6}{7.n^2} + \frac{1}{n} \longrightarrow 0$ as $n \longrightarrow \infty$

Since $\frac{B_n}{n^2} \longrightarrow 0$ as $n \longrightarrow \infty$, $WLLN$ holds for the given sequence.

P3:

Show that $WLLN$ holds to the mean of a sequence of independent r.v s X_k specified by

$$P\left(X_k = \pm\sqrt{ln\, k}\right) = \frac{1}{2}$$

P3:

Show that $\boldsymbol{WLLN}$ holds to the mean of a sequence of independent r.v s $\boldsymbol{X_k}$ specified by

$$\boldsymbol{P\left(X_k = \pm\sqrt{\ln k}\right) = \frac{1}{2}}$$

Solution:

Here $E(X_k) = \sqrt{ln\,k}\,.\frac{1}{2} - \sqrt{ln\,k}\,,\frac{1}{2} = 0$ and

$$V(X_k) = E(X_k^2) = \frac{1}{2}ln\,k + \frac{1}{2}ln\,k = ln\,k$$

Let $S_n = \sum_{k=1}^{n} S_k$. Then

$V(S_n) = \sum_{k=1}^{n} V(X_k)$ $\quad\quad$ ($\because X_K\,'s$ are independent)

$$= \sum_{k=1}^{n} \ln k = ln\,1 + ln\,2 + \cdots + ln\,n = ln(n!)$$

$\Longrightarrow V(S_n) = B_n = ln(n!)$ ... (1)

By Stirling's approximation

$$n! = \sqrt{2\pi}e^{-n}\,n^{n+\frac{1}{2}} \quad \text{... (2)}$$

We have $\frac{B_n}{n^2} = \frac{ln\left(\sqrt{2\pi}e^{-n}.n^{n+\frac{1}{2}}\right)}{n^2} = \frac{-n+\left(n+\frac{1}{2}\right)\ln n + ln\sqrt{2\pi}}{n^2}$

$$= \frac{\left(ln\sqrt{2\pi} - n\right)}{n^2} + \frac{1}{n}\left(1 + \frac{1}{2n}\right)\ln n$$

$$= \frac{\ln\sqrt{2\pi} - n}{n^2} + \left(1 + \frac{1}{2n}\right)\ln\left(n^{1/n}\right) \longrightarrow 0 \text{ as } n \longrightarrow \infty \quad \left(\because \lim_{n\to\infty} n^{1/n} = 1\right)$$

Hence $\lim_{n\to\infty} \frac{B_n}{n^2} = 0$. Thus, the sequence $\{X_n\}$ holds $WLLN$.

P4:

The variates $X_1, X_2, ..$ have equal expectations and finite variations. Is $WLLN$ applicable to this sequence if all covariances σ_{ij} are negative?

P4:

The variates $X_1, X_2, ..$ have equal expectations and finite variations. Is $WLLN$ applicable to this sequence if all covariances σ_{ij} are negative?

Solution:

We have

$$V(aX + bY) = a^2V(X) + b^2V(Y) + 2ab\ cov\ (X, Y)$$

$$= a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\sigma_{XY}$$

$$\leq a^2\sigma_X^2 + b^2\sigma_Y^2 \qquad (\because\ \sigma_{XY} < 0)$$

$$\Rightarrow 0 \leq V(aX + bY) \leq a^2\sigma_X^2 + b^2\sigma_Y^2$$

$$\therefore\ 0 \leq V\left(\frac{X_1 + \ldots + X_n}{n}\right) \leq \frac{1}{n^2} \sum_{i=1}^{n} V\left(X_i\right) < \frac{A}{n} \qquad \ldots(1)$$

where A is the upper bound of $V(X_i)\ \forall\ i = 1, 2, \ldots . n$

Now, $\frac{B_n}{n^2} = V\left(\frac{X_1 + \cdots + X_n}{n}\right) < \frac{A}{n} \longrightarrow 0$ as $n \longrightarrow \infty$. (from (1))

Thus, $WLLN$ is applicable to this sequence.

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

$$\sigma_n{}^2 = V(X_n) = E\left(X_n{}^2\right) = 2^{2n} \frac{1}{2^{2n+1}} + 2^{2n} \frac{1}{2^{2n+1}} = 1$$

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

$$\sigma_k{}^2 = V(X_k) = E\left(X_k{}^2\right) = k^{2\alpha} \frac{1}{2} + k^{2\alpha} \frac{1}{2} = k^{2\alpha}$$

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

$$E(S_n) = nE({X_1}^2) = n\left[V(X_1) + \left(E(X_1)\right)^2\right] = n(\sigma^2 + \mu^2)$$

$$\Rightarrow E(S_n) = n(\sigma^2 + \mu^2)$$

$$\Rightarrow E\left(\frac{S_n}{n}\right) = \sigma^2 + \mu^2$$

By Theorem 2,

$$\frac{S_n}{n} \xrightarrow{a.s} E\left(\frac{S_n}{n}\right) = (\sigma^2 + \mu^2)$$

$$\Rightarrow \frac{{X_1}^2 + \cdots + {X_n}^2}{n} \xrightarrow{a.s} c, \text{ where } c = \sigma^2 + \mu^2 .$$

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 $\{X_n\}$.

4.4. Strong Law of Large Numbers

Exercise:

For the following sequences of independent r.vs, does the SLLN hold or not?

1. $P(X_k = \pm 2^{k+1}) = 2^{-(k+3)}\, , P(X_k = 0) = 1 - 2^{-(k+2)}$

2. $P(X_k = \pm 2^{-k}) = \frac{1}{2^{k+1}}\, , P(X_k = \pm 1) = \frac{1}{2}(1 - 2^{-k})$

3. Examine if SLLN holds for the sequence $\{X_i\}$ of i.i.d.r.vs with
$$P[X_i = (-1)^{k-1}k] = \frac{6}{\pi^2 k^2}\, , k = 1,2,3, \dots\ i = 1,2, \dots$$

4. Let $X_1\, , X_2, \dots, X_n$ be jointly normal with $E(X_i) = 0\, , E\left({X_i}^2\right) = 1$ for all i and
$$cov\left(X_i, X_j\right) = \begin{cases} \rho\ , & |j - i| = 1 \\ 0, & \text{otherwise} \end{cases}$$

Examine if SLLN holds for the sequence $\{X_i\}$.

Answer

1. No
2. Yes
3. Yes
4. Yes

P1:

Investigate the a.s. convergence of $\{\overline{X_n}\}$ to $\mathbf{0}$ where X_ns are independent and

$$P(X_n = 0) = 1 - \frac{1}{n}, P(X_n = \pm 1) = \frac{1}{2n}$$

P1:

Investigate the a.s. convergence of $\{\overline{X_n}\}$ to 0 where X_ns are independent and

$$P(X_n = 0) = 1 - \frac{1}{n}, P(X_n = \pm 1) = \frac{1}{2n}$$

Solution:

Here $E(X_n) = 0\left(1 - \frac{1}{n}\right) + \frac{1}{2n} - \frac{1}{2n} = 0$ and

$\sigma_n^2 = V(X_n) = E(X_n^2) = 0^2\left(1 - \frac{1}{n}\right) + 1^2.\frac{1}{2n} + 1^2.\frac{1}{2n} = \frac{1}{n}$

Thus, we have

$\sum_{n=1}^{\infty} \frac{\sigma_n^2}{n^2} = \sum_{n=1}^{\infty} \frac{1}{n}\frac{1}{n^2} = \sum_{n=1}^{\infty} \frac{1}{n^3}$ is convergent and by $SLLN$ $\frac{S_n}{n} \xrightarrow{a.s} E\left(\frac{S_n}{n}\right)$

But $E\left(\frac{S_n}{n}\right) = \frac{1}{n}E\left(\sum_{k=1}^{n} X_i\right) = \frac{1}{n}\sum_{k=1}^{n}(X_i) = 0$ and $\frac{S_n}{n} = \bar{X}_n$

Therefore, $\bar{X}_n \xrightarrow{a.s.} 0$

P2:

A variate X_k has the distribution

$$P(X_k = 0) = 1 - \left(\frac{2}{3^{2k+2}}\right),\ P\left(X_k = \pm 3^k\right) = 3^{-(2k+2)}$$

If $\{X_k\}$ is a sequence of independent r.vs, then show that $\{X_k\}$ obeys $SLLN$.

P2:

A variate $\boldsymbol{X_k}$ has the distribution

$$\boldsymbol{P(X_k = 0) = 1 - \left(\frac{2}{3^{2k+2}}\right), P\left(X_k = \pm 3^k\right) = 3^{-(2k+2)}}$$

If $\{X_k\}$ is a sequence of independent r.vs, then show that $\{X_k\}$ obeys $SLLN$.

Solution:

Here $E(X_k) = 3^k \; 3^{-(2k+2)} - 3^k \; 3^{-(2k+2)} = 0$ and

$\sigma_k^2 = V(X_K) = E(X_n^2) = 3^{2k} 3^{-(2k+2)} + 3^{2k} \; 3^{-(2k+2)} = \frac{2}{9}$

Then we have

$\sum_{k=1}^{\infty} \frac{\sigma_k^2}{k^2} \frac{1}{k^2} = \frac{2}{9} \sum_{k=1}^{\infty} \frac{1}{k^2}$ Converges.

Thus $\{X_n\}$ obeys the $SLLN$.

P3:

Let $\{X_n\}$ be a sequence of independent r.vs such that

$$\boldsymbol{P(X_n = \pm 1) = \frac{1}{2}(1 - 2^{-n}), P(X_n = \pm 2^{-n}) = 2^{-n-1}}$$

Does the $SLLN$ hold for this sequence?

P3:

Let $\{X_n\}$ be a sequence of independent r.vs such that

$$P(X_n = \pm 1) = \frac{1}{2}(1 - 2^{-n}), P(X_n = \pm 2^{-n}) = 2^{-n-1}$$

Does the $SLLN$ hold for this sequence?

Solution:

Here $E(X_n) = 1\frac{1}{2}(1 - 2^{-n}) - 1\,\frac{1}{2}(1 - 2^{-n}) + 2^{-n}2^{-n-1} - 2^{-n}2^{-n-1} = 0$

and $\sigma_n^2 = V(X_n) = E(X_n^2) = 1^2\frac{1}{2}(1 - 2^{-n}) + 1^2\frac{1}{2}(1 - 2^{-n}) + 2^{-2n}2^{-2n-1}$

$+2^{-2n}\,2^{-n-1}$

$$= 2\,\frac{1}{2}(1 - 2^{-n}) + 2.\,2^{-2n}2^{-n-1}$$

$\Longrightarrow \sigma_n^2 = 1 - 2^{-n} + 2^{-3n}$

Further, we have $\sum\limits_{n=1}^{\infty}\frac{\sigma_n^2}{n^2} = \sum\limits_{n=1}^{\infty}\frac{1}{n^2} - \sum\limits_{n=1}^{\infty}\frac{1}{2^n n^2} + \sum\limits_{n=1}^{\infty}\frac{1}{2^{3n} n^2}$ converges.

Thus $\{X_n\}$ obeys $SLLN$.

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 $\epsilon > 0$

$$P(\overline{X_n} \geq \epsilon) = \frac{\sigma}{\epsilon\sqrt{n}} \frac{1}{\sqrt{2\pi}} e^{-\frac{n\,\epsilon^2}{2\sigma^2}} \text{ as } n \longrightarrow \infty.$$

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, $S_n \sim AN\left(E(S_n), V(S_n)\right)$

$$\Rightarrow \; S_n \sim AN\left(np, np(1-p)\right) \qquad \ldots (1)$$

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, $\lim\limits_{n\to\infty} P\left(a \le \frac{S_n - np}{\sqrt{npq}} \le b\right) = P(a \le Z \le b) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}z^2} dz$

and the RHS can be evaluated using standard normal tables for given real numbers a and b.

(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 $V(S_n) = V\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} V(X_i) \qquad (\because \; X_i s \text{ are independent })$

$$= \sum_{i=1}^{n} rp(1-p) = nrp(1-p)$$

$$\left(\because \quad V(X_i) = rp(1-p)\right)$$

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 $i = 1, 2, \ldots n$

$$\therefore\ E(S_n) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} \lambda = n\lambda \quad \text{and}$$

$$V(S_n) = V\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} V(X_i) = n\lambda$$

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 $Z \sim N(0, 1)$

$$= P(-2.45 \le Z \le 0.82)$$

$$= P(-2.45 \le Z \le 0) + P(0 \le Z \le 0.82)$$

$$= P(0 \le Z \le 2.45) + P(0 \le Z \le 0.82)$$

$$= 0.4929 + 0.2938 \quad \text{(From standard normal table)}$$

$$= 0.7868$$

4.5. Central Limit Theorem

Exercise:

1. The lifetime of a certain brand of an electric bulb may be considered as a r.v. with mean $1200\ hrs$ and standard deviation $250\ hrs$. Find the probability, using CLT that the average lifetime of 60 bulbs exceeds 1250 hrs.

2. A distribution has unknown mean μ and variance σ^2 equal to 1.5. Use CLT to find how large a sample should be taken from the distribution in order that the probability will be at least 0.95 that the sample mean will be within 0.5 of the population mean.

3. A random sample of size 100 is taken from a population whose mean is 60 and variance is 400. Use CLT, with what probability can we assert that the mean of the sample will not differ from $\mu = 60$ by more than 47.

4. The guaranteed average life of a certain type of electric light bulb is $1000\ hrs$ with a standard deviation of $125 hrs$. It is decided to sample the output so as to ensure that 90% of the bulbs do not fall short of the guaranteed average by more than 2.5%. Use CLT to find the minimum sample size.

5. If X_i , $i = 1,2, \dots ,50$ are independent r.vs, each having a Poisson distribution with parameter $\lambda = 0.03$ and $S_n = X_1 + \cdots + X_n$, evaluate $P(S_n \geq 3)$, using CLT.

Answers:

1. 0.0606
2. at least 24
3. 0.9544
4. 41
5. 0.1112

P1:

Show that WLLN follows from CLT for the sequence of i.i.d.r.vs .

P1:

Show that WLLN follows from CLT for the sequence of i.i.d.r.vs .

Solution:

Let $\{X_n\}$ be a sequence of i.i.d.r.vs with $E(X_i) = \mu$ and $V(X_i) = \sigma^2 \; \forall i$.

Then $\{X_n\}$ obeys WLLN if

$$\lim_{n \longrightarrow \infty} P\left[\left|\frac{S_n}{n} - \mu\right| > \epsilon\right] = 0 \text{ (By Khinchine's WLLN)}$$

for $\epsilon > 0$ and $S_n = \sum_{i=1}^{n} X_i$.

$$\Longrightarrow \lim_{n \longrightarrow \infty} P\left[\left|\frac{S_n}{n} - \mu\right| < \epsilon\right] = 1 \qquad .... \; (1)$$

Since $\epsilon > 0$ is fixed, for any positive constant k, we have

$$k\sigma\sqrt{n} < \epsilon n \qquad .... \; (2)$$

for all sufficiently large values of n (because any large multiple of $\sqrt{n}$ is negligible in comparison with any small multiple of n)

From (2), we have $|S_n - n\mu| < k\sigma\sqrt{n} \Longrightarrow \{|S_n - n\mu| < \epsilon n\}$

$i.e., \frac{|S_n - n\mu|}{\sigma\sqrt{n}} < k \Longrightarrow \left|\frac{S_n - n\mu}{n}\right| < \epsilon$

Therefore, $P\left\{\frac{|S_n - n\mu|}{\sigma\sqrt{n}} < k\right\} \leq P\left\{\left|\frac{S_n - n\mu}{n}\right| < \epsilon\right\}$

$$i.e, \lim_{n \longrightarrow \infty} P\left\{\frac{|S_n - n\mu|}{\sigma\sqrt{n}} < k\right\} \leq \lim_{n \longrightarrow \infty} P\left\{\left|\frac{S_n - n\mu}{n}\right| < \epsilon\right\} \qquad ... \; (3)$$

As $n \longrightarrow \infty$, by Lindeberg-Levy CLT, $Z = \frac{S_n - n\mu}{\sigma\sqrt{n}} \sim N(0,1)$ $i.e.$, standard normal variate. Then, we can choose large k such that

$$\lim_{n \longrightarrow \infty} P\left\{\frac{|S_n - n\mu|}{\sigma\sqrt{n}} < k\right\} = P(|Z| < k) = 1 \qquad ...(4)$$

From (3) and (4), we have

$$\lim_{n \longrightarrow \infty} P\left\{\left|\frac{S_n}{n} - \mu\right| < \epsilon\right\} \geq 1$$

$\Longrightarrow \lim_{n \longrightarrow \infty} P\left\{\left|\frac{S_n}{n} - \mu\right| < \epsilon\right\} = 1$ ($\because$ Probability can't be more than 1) which is 1.

Therefore, WLLN holds.

P2:

Examine if CLT holds for the sequence $\{X_k\}$ with p.m.f. $P\left(X_k = \pm k^{\lambda}\right) = \frac{1}{2}$.

P2:

Examine if CLT holds for the sequence $\{X_k\}$ with p.m.f. $P(X_k = \pm k^{\lambda}) = \frac{1}{2}$.

Solution:

We have $\mu_k = E(X_k) = k^{\lambda}\frac{1}{2} - k^{\lambda}\frac{1}{2} = 0$,

$\sigma_k{}^2 = V(X_k) = k^{2\lambda}\frac{1}{2} + k^{2\lambda}\frac{1}{2} = k^{2\lambda}$ and

$\rho_k^3 = E\{|X_k - 0|^3\} = E(|X_k|^3) = k^{3\lambda}\cdot\frac{1}{2} + k^{3\lambda}\cdot\frac{1}{2} = k^{3\lambda}$

Let $S_n = \sum_{k=1}^{n} X_k$. Then we have $\mu = \sum_{k=1}^{n} \mu_k = 0$, $\sigma^2 = \sum_{k=1}^{n} \sigma_k{}^2 = \sum_{k=1}^{n} k^{2\lambda}$

$\rho^3 = \sum_{k=1}^{n} \rho_k{}^3 = \sum_{k=1}^{n} k^{3\lambda}$ and

$$\frac{\rho^3}{\left(\sigma^2\right)^{\frac{3}{2}}} = \frac{\sum_{k=1}^{n} k^{3\lambda}}{\left(\sum_{k=1}^{n} k^{2\lambda}\right)^{\frac{3}{2}}} = \frac{n^{3\lambda+1}}{3\lambda+1} \times \left(\frac{2\lambda+1}{n^{2\lambda+1}}\right)^{\frac{3}{2}} \qquad \left(\begin{array}{l} \because \sum_{k=1}^{n} k^{\alpha} = \int_0^n x^{\alpha}dx = \dfrac{n^{\alpha+1}}{\alpha+1} \\ Euler - maclaurian\ formula \end{array}\right)$$

$$= \frac{\left(2\lambda+1\right)^{\frac{3}{2}}}{\left(3\lambda+1\right)} n^{(3\lambda+1)-(2\lambda+1)\frac{3}{2}}$$

$$= \frac{\left(2\lambda+1\right)^{\frac{3}{2}}}{\left(3\lambda+1\right)} n^{-\frac{1}{2}} \rightarrow 0 \ as\ n \rightarrow \infty$$

Since Liapnounoff's condition is satisfied, CLT holds.

P3:

Examine if CLT holds for the sequence $\{X_k\}$ with p.m.f. $P\left(X_k = \pm 2^{-k}\right) = \frac{1}{2}$.

P3:

Examine if CLT holds for the sequence $\{X_k\}$ with p.m.f. $P\left(X_k = \pm 2^{-k}\right) = \frac{1}{2}$.

Solution:

We have $\mu_k = E(X_k) = 2^{-k} \cdot \frac{1}{2} - 2^{-k} \cdot \frac{1}{2} = 0$,

$\sigma_k{}^2 = V(X_k) = E\left(X_k{}^2\right) = 2^{-2k} \cdot \frac{1}{2} + 2^{-2k} \cdot \frac{1}{2} = 2^{-2k}$ and

$\rho_k{}^3 = E\{|X_k - 0|^3\} = E(|X_k|^3) = 2^{-3k} \cdot \frac{1}{2} + 2^{-3k} \cdot \frac{1}{2} = 2^{-3k}$

Let $S_n = \sum_{k=1}^{n} X_k$. Then we have $\mu = \sum_{k=1}^{n} \mu_k = 0$

$$\sigma^2 = \sum_{k=1}^{n} \sigma_k{}^2 = \sum_{k=1}^{n} 2^{-2k} = \sum_{k=1}^{n} \frac{1}{2^{2k}} = \sum_{k=1}^{n} \frac{1}{4^k} = \frac{1}{4} + \frac{1}{4^2} + \ldots + \frac{1}{4^n} = \frac{1}{3}\left(1-4^{-n}\right)$$ and

$$\rho^3 = \sum_{k=1}^{n} \rho_k{}^3 = \sum_{k=1}^{n} 2^{-3k} = \sum_{k=1}^{n} \frac{1}{8^k} = \frac{1}{8} + \frac{1}{8^2} + \ldots + \frac{1}{8^n} = \frac{1}{7}\left(1-8^{-n}\right)$$ and

$$\frac{\rho^3}{\left(\sigma^2\right)^{\frac{3}{2}}} = \frac{\frac{1}{7}\left(1-8^{-n}\right)}{\left[\frac{1}{3}\left(1-4^{-n}\right)\right]^{\frac{3}{2}}} \to \frac{1}{7}\left(\frac{1}{3}\right)^{-\frac{3}{2}} \quad as\ n \to \infty$$

Since Liapounoff's condition is not satisfied, we can't say CLT holds or not.

P4:

Let $X_1, \dots, X_n$ be i.i.d. standardized variates with $E({X_i}^n) < \infty$. Find the limiting distribution of

$$Z_n = \frac{\sqrt{n}(X_1X_2 + X_3X_4 + \cdots + X_{2n-1}X_{2n})}{\left({X_1}^2 + {X_2}^2 + \cdots + {X_{2n}}^2\right)}$$

P4:

Let $X_1, \dots, X_n$ be i.i.d. standardized variates with $E(X_i^n) < \infty$. Find the limiting distribution of

$$\boldsymbol{Z_n = \frac{\sqrt{n}(X_1X_2 + X_3X_4 + \cdots + X_{2n-1}X_{2n})}{\left(X_1{}^2 + X_2{}^2 + \cdots + X_{2n}{}^2\right)}}$$

Solution:

Since X_is are i.i.d standardized variates, we have

$$E(X_i) = 0\,, V(X_i) = E\left(X_i{}^2\right) = 1\,, i = 1,2, \dots, n$$

Let $Y_i = X_{2i-1}X_{2i}\;\;, i = 1,2, \dots, n$

$\Longrightarrow E(Y_i) = E(X_{2i-1})E(X_{2i}) = 0$ (∵ X_is are independent)

and $V(Y_i) = E\left(Y_i{}^2\right) = E\left[X_{2i-1}{}^2\, X_{2i}{}^2\right] = E\left[X_{2i-1}{}^2\right] E\left[X_{2i}{}^2\right] = 1 \cdot 1 = 1$

Hence $Y_i\,, i = 1,2, \dots, n$ are also i.i.d.r.vs. Hence CLT holds for $S_n = \sum_{i=1}^{n} Y_i$. Further,

$$E(S_n) = \sum_{i=1}^{n} E(Y_i) = 0 \text{ and}$$

$$V(S_n) = \sum_{i=1}^{n} V(Y_i) = n$$

Then by CLT

$$U_n = \frac{S_n - E(S_n)}{\sqrt{V(S_n)}} = \frac{X_1X_2 + \cdots + X_{2n-1}X_{2n}}{\sqrt{n}} \xrightarrow{L} Z \text{ where } Z \sim N(0,1) \text{ as } n \longrightarrow \infty.$$

Also $E\left(X_i{}^2\right) = 1$ (finite), $i = 1,2, \dots, n$.

Hence, by Khinchine's theorem, WLLN applies to the sequence $\left\{X_i{}^2\right\}$, $i = 1,2, \dots, 2n$ so that

$$V_n = \frac{X_1{}^2 + X_2{}^2 + \cdots + X_{2n}{}^2}{2n} \xrightarrow{P} E\left(X_i{}^2\right) = 1 \text{ as } n \longrightarrow \infty.$$

Hence, by Slutsky's theorem, we have

$$\lim_{n\to\infty}(U_n \cdot V_n) = \frac{2\sqrt{n}(X_1X_2 + \cdots + X_{2n-1}X_{2n})}{{X_1}^2 + \cdots + {X_n}^2} \xrightarrow{L} \frac{Z}{1} = Z \sim N(0,1)$$

Thus, $\frac{\sqrt{n}(X_1X_2+\cdots+X_{2n-1}X_{2n})}{{X_1}^2+\cdots+{X_n}^2} \xrightarrow{L} \frac{Z}{2} \sim N\left(0, \frac{1}{4}\right)$ (If $Z \sim N(0,1)$, then $cZ \sim N(0, c^2)$).

UNIT 16 BASIC CONCEPTS OF SAMPLING

Structure

16.0 OBJECTIVES

After going through this unit you should be able to:

- explain the concepts of population, sample, parameter, statistic, estimator and estimate;
- distinguish between a census and a sample survey;
- explain the advantages of a sample survey;
- distinguish between sampling error and non-sampling error;
- explain the concept of sampling distribution; and
- explain the concept of standard error.

16.1 INTRODUCTION

We need data for the construction of national income accounts, input-output tables, various production indices, price indices and a host of other quantitative indicators. It is very clear that without the relevant data, we will not be able to formulate policy objectives for a complex economy like ours. In a sense, modern society is increasingly becoming an information society. In this society, various economic and social processes are represented by certain quantitative characteristics that require various kinds of information in the form of data.

The task of collecting data is getting increasingly complex and difficult. The total number of units to be consulted and investigated for the required information may be too large and our resources in terms of money, time or personnel may be limited. Moreover, obtaining error-free information from such a large-scale investigation makes the job even more daunting. As a result, very often we try to obtain the required information from a smaller group that is easier to handle and control. Here, however, it is important to ensure that this smaller group is truly representative of the entire collection of relevant units. The subject matter of sampling provides a mathematical theory for obtaining such kind of a representative group.

16.2 CENSUS AND SAMPLE SURVEY

In this Section, we will distinguish between the census and sampling methods of collecting data. We will try to explain the meaning and coverage of census survey and sample survey.

16.2.1 Population and Census

We have a collection of units relevant for a particular enquiry. A unit, in this connection, is an entity on which we can make observations according to a well-defined procedure. The entire collection of such units is called a *population* or *universe*. Thus, we may have a population of human beings, cattle, trees, prices, production, etc.

You can make out that a population can be finite or infinite. If the number of units is finite, it is a finite population and if the number of units is infinite, it is an example of an infinite population. Usually in practice, we are concerned with a finite population.

When an inquiry is based upon obtaining information from all the units of a population, the procedure is known as the *complete enumeration method* or the *census method*.

16.2.2 Sample and Sample Survey

When we have a collection of a part or section of the population, it is called a *sample*. A census, as we have seen earlier, is based upon obtaining information from every member of the population. However, in order to obtain information about certain characteristic of the population, we need not always resort to a census. In practice, we get quite satisfactory results by studying an appropriate sample from the population. The procedure of obtaining a sample is known as *sample survey*. In the case of a census, we examine the entire population; on the other hand, when we take a sample, we consider a representative fraction of the population and use the sample information to infer about the entire population.

16.3 SOME CONCEPTS

We explain below some of the concepts frequently used in sampling theory.

16.3.1 Parameter

In a statistical inquiry, our interest lies in one or more characteristics of the population. A measure of such a characteristic is called a *parameter*. For example, we may be interested in the mean income of the people of some region for a particular year. We may also like to know the standard deviation of these incomes of the people. Here, both mean and standard deviation are parameters.

Parameters are conventionally denoted by Greek alphabets. For example, the population mean can be denoted by μ and population standard deviation can be denoted by σ.

It is important to note that the value of a parameter is computed from all the population observations. Thus, the parameter 'mean income' is calculated from all the income figures of different individuals that constitute the population. Similarly, for the calculation of the parameter 'correlation coefficient of heights and weights', we require the values of all the pairs of heights and weights in a population. Thus, we can define a *parameter as a function of the population values.* If θ is a parameter that we want to obtain from the population values $X_1, X_2, \cdots X_N$, then

$$\theta = f(X_1, X_2, \cdots X_N)$$

16.3.2 Statistic

While discussing the census and the sample survey, we have seen that due to various constraints, sometimes it is difficult to obtain information about the whole population. In other words, it may not be always possible to compute a population parameter. In such situations, we try to get some idea about the parameter from the information obtained from a sample drawn from the population. This sample information is summarised in the form of a *statistic*. For example, sample mean or sample median or sample mode is called a statistic. Thus, a statistic is calculated from the values of the units that are included in the sample. So, a *statistic can be defined as a function of the sample values.* Conventionally, a statistic is denoted by an English alphabet. For example, the sample mean may be denoted by $\bar{x}$ and the sample standard deviation may be denoted by *s*. If *T* is a statistic that we want to obtain from the sample values $x_1, x_2, \cdots x_n$, then

$$T = f(x_1, x_2, \cdots x_n)$$

16.3.3 Estimator and Estimate

The basic purpose of a statistic is to estimate some population parameter. The procedure followed or the formula used to compute a statistic is called an *estimator* and the value of a statistic so computed is known as an *estimate.*

If we use the formula $\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$ for calculating a statistic, then this formula is an estimator. Next, if we use this formula and get $\bar{x} = 10$, then this '10' is an estimate.

16.4 NON-SAMPLING AND SAMPLING ERRORS

As mentioned above the basic purpose of sampling is to draw inferences about the population on the basis of the sample. For example, we have to find out the per capita income of a village. Due to shortage of time, money and personnel we do not undertake a complete census and opt for a sample survey. In this case it is very likely that the per capita income obtained from the sample is not equal to the actual per capita income of the village. This discrepancy could arise because of two reasons:

i) Since we are collecting data from only a part of the population (i.e., the sample selected by us), sample mean (per capita income in this case) is not equal to population mean. If at all both are equal, it is a rare coincidence! If we take sample mean as population mean we are committing an error called sampling error.

ii) A second source of error could arise because of wrong reporting or recording or tabulation or processing of data. This type of error is termed non-sampling error. Remember that non-sampling error, as its name suggests, has nothing to do with our sampling process. Wrong reporting or recording or processing of data can take place in a sample survey also.

We explain the sources of these errors below.

16.4.1 Non-Sampling Error

Various sources of non-sampling error are given below:

1) **Error due to measurement**

It is a well-known fact that precise measurement of any magnitude is not possible. If some individuals, for example, are asked to measure the length of a particular piece of cloth independently up to, say, two decimal points; we can be quite sure that their answers will not be the same. In fact, the measuring instrument itself may not have the same degree of accuracy.

In the context of sampling the respondents of an inquiry, for example, may not be able to provide the accurate data about their incomes. This may not be a problem with individuals earning fixed incomes in the form of wages and salaries. However, self-employed persons may not be able to do so.

2) **Error due to non-response**

Sometimes the required data are collected by mailing questionnaires to the respondents. Many of such respondents may return the questionnaires with incomplete answers or may not return them at all. This kind of an attitude may be due to:

a) the respondents are too casual to fill up the answers to the questions asked

b) they are not in a position to understand the questions, or

c) they may not like to disclose the information that has been sought.

We should note that the error due to non-response may also arise because of the possibility of the questionnaire being lost in transit.

If the data are collected through personal interviews, some of the reasons for the error due to non-response pointed out above may not arise. However, in that case this error may arise because some of the individuals:

a) may not like to give the information, or

b) may not simply be available even after repeated visits.

3) **Error in recording**

This type of error may arise at the stage when the investigator records the answers or even at the tabulation stage. A major reason for such error is the carelessness on the part of the investigator.

4) **Error due to inherent bias of the investigator**

Every individual suffers from personal prejudices and biases. Despite the provision of the best possible training to the investigators, their personal biases may come into play when they interpret the questions to be put to the respondents or record the answers to these questions.

In complete enumeration the extent of non-sampling error tends to be significantly large because, generally, a large number of individuals are involved in the data collection process. We try to minimise this error through:

i) a careful planning of the survey,

ii) providing proper training to the investigators,

iii) making the questionnaire simple.

However, we would like to emphasize that complete enumeration is always prone to large non-sampling errors.

16.4.2 Sampling Error

By now it should be clear that in the sampling method also, non-sampling error may be committed. It is almost impossible to make the data absolutely free of such errors. However, since the number of respondents in a sample survey is much smaller than in census, the non-sampling error is generally less pronounced in the sampling method. Besides the non-sampling errors, there is sampling error in a sample survey. Sampling error is the absolute difference between the parameter and the corresponding statistic, that is, $|T-\theta|$.

Sampling error is not due to any lapse on the part of the respondent or the investigator or some such reason. It arises because of the very nature of the procedure. It can never be completely eliminated. However, we have well developed sampling theories with the help of which the effect of sampling error can be minimised.

16.5 ADVANTAGES OF SAMPLE SURVEY

There are important advantages of a sample survey over complete enumeration or census method. Some of these advantages are mentioned below.

i) **Practicability**

Sometimes, a census may not be practicable due to the enormity of the task required in the collection of data of a large population. In such a situation, a sample survey may be quite practicable.

ii) **Speed**

The data may be collected and summarised faster in a sample survey than in a census. This may be an important advantage, particularly, when the information is urgently needed.

iii) **Accuracy**

In any survey, census or sample, the required information is obtained by filling in the questionnaires. It has been observed that more accurate results are achieved when the investigators themselves fill in the questionnaire instead of the respondents filling it. Again, personal interviews may result in more accurate information than sending the questionnaires to the respondents by post and requesting them to fill in these questionnaire. Normally, the number of investigators involved in an inquiry varies directly with the number of respondents covered in the inquiry. As a result, personal interviews prove to be easier in the case of a sample survey than in a census. In fact, a sample survey has a greater scope to employ more efficient and better-trained investigators. In the case of a sample survey, the investigators can devote more time to each respondent. Thus, although a sample survey can have less coverage than a census, it may have greater accuracy of the results.

iv) **Cost**

It is obvious that a sample survey results in less expenditure than a complete enumeration. After all, in a survey only part of the population is involved. The cost components of an inquiry are:

a) Overhead cost of the organisation conducting the survey,

b) Cost of collecting the data,

c) Cost of processing and tabulating the data, and

d) Cost of publication of results of the survey.

In this cost break-up, items (b) and (c) are in the nature of variable costs, whereas (a) and (d) are the fixed cost items. As a result, items (b) and (c) will definitely be much smaller in a sample survey than for a census. We should note that the designing of a proper sample survey and the selection of an appropriate sample may entail considerable expenditure. However, generally it has been observed that a sample survey is less costly than complete enumeration.

Check Your Progress 1

1) Define the following concepts:

   a) Population

   b) Sample

   c) Parameter

   d) Statistic

   ....

   ....

   ....

   ....

2) Distinguish between the following:

   a) Estimator and Estimate

   b) Census and Sample Survey

   c) Sampling error and non-sampling error

....

....

....

....

3) What are the advantages of sampling over a census?

....

....

....

....

16.6 TYPES OF SAMPLING

The method of selecting a sample from a given population is called *sampling*. Basically there are two types of sampling, viz., probability sampling and non-probability sampling. In probability sampling the sampling units are selected according to some chance mechanism or probability of selection. On the other hand, non-probability sampling is based on judgement or discretion of the person making a choice. Thus in non-probability sampling certain units may be selected because of convenience or they serve a purpose or the researcher feels that these units are representative of the population. No random selection on the basis of chance mechanism is involved here.

16.6.1 Probability Sampling

It is also called random sampling. It is a procedure in which every member of the population has a chance or probability of being selected in the sample. It is in this probabilistic sense that the sample is random. The word 'random' does not mean that the sample is obtained in a haphazard manner without following any rule.

Random sampling is based on the well-established principles of probability theory. There are quite a few variants of the random sampling, viz., simple random sampling, systematic random sampling and stratified random sampling. We discuss these types below.

a) Simple Random Sampling

If there is not much variation in the characteristics of the members of a population, we can follow the method of simple random sampling. In this method, we consider the population in its entirety as a homogeneous group and follow the principle of random sampling to choose the members for the sample.

There are two variants of simple random sampling, viz., simple random sampling with replacement (SRSWR) and simple random sampling without replacement (SRSWOR). This difference pertains to the way the sample units are selected. According to the procedure of simple random sampling with replacement (SRSWR), we draw one unit from the population, note down its features and put it back to the whole lot in the sense that the unit again becomes eligible for selection. In this way, the total number of units in the population always remains the same. In other words, the composition of the population remains unchanged, and each member of the population has the *same chance* or probability of being selected in the sample. In fact, if N is the size of the population, this probability is $\frac{1}{N}$.

On the other hand, in the case of simple random sampling without replacement, the unit once selected is not returned to the population in the sense that it becomes ineligible for selection again. As a result, after each successive draw, the composition of the population changes. Therefore, for subsequent draw from the population the probability of any particular unit being picked up also gets changed. Let us try to understand this. Suppose, the population size is N and we want to draw a sample of size n from it by the principle of SRSWOR. Before the first unit is drawn, each unit of the population has the *same* chance ($\frac{1}{N}$) of being selected in the sample. Once the first member of the sample is selected, each of the *remaining N-1* members of the population has an equal chance of $\frac{1}{N-1}$ of selection in the sample. Finally, before the n^{th} member of the sample is chosen, each of the *remaining* members of the population has an equal chance of

$\frac{1}{N-(n+1)} = \frac{1}{N-n+1}$ of being included in the sample.

We should note that from a population of size N, the number of samples of size n that can be drawn with replacement is N^n and the number of samples that can be drawn without replacement is ${}^N c_n$.

Example 16.1

Suppose a population consists of the following 5 units (4, 5, 7, 9, 10). How many samples of size 2 can be drawn from it?

i) If we follow the procedure of SRSWR the number of samples that can be selected is

   $= N^n = 5^2 = 25$.

   The possible samples are given by

   (4, 4), (4, 5), (4, 7), (4, 9), (4, 10), (5, 4), (5, 5), (5, 7), (5, 9), (5, 10), (7, 4), (7, 5), (7, 7), (7, 9), (7, 10), (9, 4), (9, 5), (9, 7), (9, 9), (9, 10), (10, 4), (10, 5), (10, 7), (10, 9), (10, 10).

   We should note that in sampling with replacement, the order in which the units are selected also matters. Thus, (4, 10) and (10, 4) are considered as two different samples.

ii) If we follow the procedure of SRSWOR the number of samples that can be selected is

   $$= {}^N c_n = {}^5 c_2 = \frac{5!}{2!(5-2)!} = \frac{5!}{2!\,3!} = \frac{5\times 4}{2\times 1} = \frac{20}{2} = 10.$$

   The possible samples are given by

   (4, 5), (5, 7), (7, 9), (9, 10), (4, 7), (4, 9), (4, 10), (5, 9), (5, 10), (7, 10).

   We should note that in sampling without replacement, once a member is selected, it cannot be selected again. Thus, samples like (4, 4), (5, 5) etc. cannot be selected. Similarly, if a sample like (4, 5) is selected, then another sample like (5, 4) cannot be selected.

b) **Systematic Random Sampling**

In this variant of random sampling, only the first unit of the sample is selected at random from the population. The subsequent units are then selected by following some definite rule. For example, suppose, we have to choose a sample of agricultural plots. In systematic random sampling, we begin with selecting one plot *at random* and then every 10^{th} plot may be selected.

c) **Stratified Random Sampling**

Stratified random sampling is the appropriate method if the population under consideration consists of heterogeneous units. Here, first we divide the population into certain homogeneous groups or strata. Secondly, from each stratum some units are selected by simple random sampling. Thirdly, after selecting the units from each stratum, they are mixed together to obtain the final sample.

Let us consider an example. Suppose, we want to estimate the per capita income of Delhi by a sample survey. It is common knowledge that Delhi is characterised by rich localities, middle class localities and poor localities in terms of the income groups of the people living in these localities. Now, each of these different localities can constitute a stratum from which some people may be selected by adopting simple random sampling procedure.

d) **Multi-Stage Random Sampling**

Let us consider a situation where we want to obtain information from a sample of households in a large city, say, Delhi. Sometimes, it may not be possible to directly take a sample of households because a list of all the households may not be easily obtained. In such a situation, one may resort to take samples in various stages. Generally, the city is divided into certain geographical areas for administrative purposes. These areas may be termed as city blocks. So in the first stage, some of such blocks may be selected by random sampling. In the next stage, from each of the selected blocks in the first stage, some households may be selected again by the principle of random sampling. In this way, ultimately a sample of households from a large city may be obtained. The above-mentioned example is the case of a two-stage random sampling. However, if the nature of the inquiry so demands, the method of sampling can be extended to more than two stages.

16.6.2 Non-Probability Sampling

We have considered the method of random sampling and some of its variants above. It should be clear that the basic objective of the principle of random sampling is to eliminate or at least minimise the effect of the subjective bias of the investigator in the selection of the population sample. But for certain purposes, there is a need for using discretion. For example, suppose a teacher has to choose 4 participants from a class of 30 students in a debate competition. Here, the teacher may select the top 4 debaters on the basis of her own conscious judgement about the top debaters in the class. This is an example of purposive sampling. In this method, the purpose of the sample guides the choice of certain members or units of the population.

16.6.3 Mixed Sampling

In mixed sampling, we have some features of both non-probability sampling and random sampling. Suppose, an institute has to send 5 students for managerial training in a company during the summer vacation. Initially, it may shortlist about 20 students who are considered to be suitable for the training by applying its own discretion. Then from these 20 students, 5 students may finally be selected by random sampling.

16.7 SAMPLING DISTRIBUTIUON

By now it should be clear that generally the size of a sample is much smaller than the parent population. Consequently, many samples can be selected from the same population which are different from one another. Since an estimate of a parameter depends upon the sample values, and these values may change from one sample to another, there can be different estimates or values of a statistic for the same parameter. This variation in values is called *sampling fluctuation*. Suppose, a number of samples, each of size n, are drawn from a population of size N and for each sample, the value of the statistic is computed. If the number of samples is large, these values can be arranged in the form of a relative frequency distribution. When the number of samples tends to infinity, the resultant relative frequency distribution of the values of a statistic is called the *sampling distribution* of the given statistic.

Suppose, we are interested in estimating the population mean (which is a parameter), denoted by μ. A random sample of size n is drawn from this population (of size N). The sample mean $\bar{x} = \frac{1}{n}\Sigma x_i$ is a statistic corresponding to the population mean μ. We should note that $\bar{x}$ is a random variable as its value changes from one sample to another in a probabilistic manner.

Example 16.2

Consider a population consisting of the following 5 units: 2, 4, 6, 8, and 10. Suppose, a sample of size 2 is to be selected from it by the method of simple random sampling without replacement. We want to obtain the sampling distribution of the sample mean and its standard error.

- The number of samples that can be selected without replacement.

$$= {}^N c_n = {}^5 c_2 = \frac{5!}{2!(5-2)!} = \frac{5!}{2!\,3!} = \frac{5\times 4}{2\times 1} = \frac{20}{2} = 10.$$

The possible samples along with the corresponding sample means $(\bar{x})$ are presented in Table 16.1.

Table 16.1 : Possible Samples and Sample Means

Sample	Sample Mean ($\bar{x}$)
(2, 4)	3
(2, 6)	4
(2, 8)	5
(2, 10)	6
(4, 6)	5
(4, 8)	6
(4, 10)	7
(6, 8)	7
(6, 10)	8
(8, 10)	9

Now, we can have a frequency distribution of the sample means:

Table 16.2: Frequency Distribution of Sample Means

Sample Mean	Frequency
($\bar{x}$)	(f)
3	1
4	1
5	2
6	2
7	2
8	1
9	1

From the frequency distribution given in Table 16.2, we can present the probability distribution of the sample mean as given in Table 16.3.

Table 16.3: Sampling Distribution of Sample Means

Sample Mean ($\bar{x}$)	Probability $\left(\frac{f}{\Sigma f}\right)$
3	$\frac{1}{10}$
4	$\frac{1}{10}$
5	$\frac{2}{10}$
6	$\frac{2}{10}$
7	$\frac{2}{10}$
8	$\frac{1}{10}$
9	$\frac{1}{10}$

We note here that $\sum f$, which, from the frequency distribution of the sample mean presented earlier, is equal to 10. In Table 16.3, we have used the relative frequency for the calculation of the probabilities.

16.8 STANDARD ERROR OF A STATISTIC

In the previous Section we learnt that we can draw a number of samples depending upon the population and sample sizes. From each sample we get a different value for the statistic we are looking for. These values can be arranged in the form of a probability distribution, which is called the sampling distribution of the concerned statistic. The statistic is also similar to a random variable since a probability is attached to each value it takes. In Table 16.3 in the previous Section we have presented the statistic along with its probability.

We have learnt in Unit 14 that mathematical expectation of a random variable is equal to its arithmetic mean. Let us find out the mathematical expectation and standard deviation of the sampling distribution.

We notice two important properties of the sampling distribution.

1) The expectation of the sampling distribution of the statistic is equal to the population parameter. Thus if we have the sampling distribution of sample means, then its expected value is equal to population mean. Symbolically, $E(\bar{x}) = \mu$.

2) The standard deviation of the sampling distribution is called 'standard error' of the concerned statistic. Thus if we have sampling distribution of sample means, then its standard deviation is called the 'standard error of sample means'. Thus standard error indicates the spread of the sample means away from the population mean. In Block 7 we would see that standard error is used for hypothesis testing and statistical estimation.

Example 16.3

Find out the standard error of the sampling distribution given in Table 16.3

We know that standard error of the sample mean is standard deviation of the sampling distribution. Thus,

$$\sigma_{\bar{x}} = \sqrt{E(\bar{x})^2 - [E(\bar{x})]^2}$$

Now,

$$E(\bar{x}) = 3\times\frac{1}{10} + 4\times\frac{1}{10} + 5\times\frac{2}{10} + 6\times\frac{2}{10} + 7\times\frac{2}{10} + 8\times\frac{1}{10} + 9\times\frac{1}{10} = \frac{60}{10} = 6$$

and

$$E(\bar{x})^2 = 9\times\frac{1}{10} + 16\times\frac{1}{10} + 25\times\frac{2}{10} + 36\times\frac{2}{10} + 49\times\frac{2}{10} + 64\times\frac{1}{10} + 81\times\frac{1}{10} = \frac{390}{10}$$
$$= 39.$$

$$\therefore \sqrt{E(\bar{x})^2 - [E(\bar{x})]^2} = \sqrt{39-36} = \sqrt{3} = 1.73.$$

Thus, the standard error of the sample mean in this case is 1.73.

Now a question may be shaping up in your mind.

Do we have to draw all possible samples to find out standard error? In Example 16.3 above we first noted down all the possible samples, arranged these in a relative frequency distribution form and thereafter calculated the standard deviation. In Example 16.3 the population size and sample size were quite small, and thus the task was manageable. But, can you imagine what would happen when we have much larger population and sample sizes? It is too difficult and cumbersome a task. In fact the entire advantages of sampling disappears if we start selecting all possible samples!

Secondly, is it possible to fit a theoretical probability distribution (discussed in Block 5) to the sampling distribution? In fact, the *Central Limit Theorem* says that, "if samples of size *n* are drawn from any population, the sample means are approximately normally distributed for large values of *n*". Thus whatever be the distribution of the population, the sampling distribution of $\bar{x}$ will be approximately normal for large enough sample sizes. If the population is normal, then sampling distribution of $\bar{x}$ is normal for any sample size. If population is approximately normally distributed than sampling distribution of $\bar{x}$ is approximately normal even for small sample size. Moreover, even if population is *not* normally distributed, sampling distribution of $\bar{x}$ is approximately normal for large sample sizes.

Thirdly, what is the relationship between standard deviation of the population from which the sample is drawn and the standard error of $\bar{x}$? Obviously, the spread of $\bar{x}$ will be less than the spread of the population units. The standard error of $\bar{x}$ is given by

$\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$ where $\sigma_{\bar{x}}$ is standard error of $\bar{x}$ and σ is standard deviation of the original population.

Thus standard error is always smaller in value than standard deviation of the population, because standard error is equal to the standard deviation of the population divided by square root of the sample size.

The above is true for simple random sampling with replacement. When sampling is without replacement in that case we have to make some finite population correction and standard error is given by $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}} \times \dfrac{N-n}{N-1}$.

When the ratio $\dfrac{n}{N}$ is very small both the procedures give almost similar results. But when sample size is not negligible compared to population size the correction factor needs to be applied.

How do we interpret the standard error? As mentioned earlier it shows the spread of the statistic. Thus, if standard error is smaller then there is a greater probability that the estimate is closer to the concerned parameter.

Example 16.4

Consider the population: 2,5,8,13

i) Calculate the population mean and the population standard deviation.

ii) Construct a sampling distribution of the sample mean when random samples of size 2 are selected from the population

   a) with replacement, and

   b) without replacement. Find the mean and the standard error of the distribution in each case.

iii) Verify that in the case of random sampling with replacement, $E(\bar{x}) = \mu$ and $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$ and in the case of random sampling without replacement, $E(\bar{x}) = \mu$ and $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}\left(\sqrt{1-\dfrac{n-1}{N-1}}\right)$.

Answer:

We have the population: 2,5,8,13; population size $N = 4$; sample size $n = 2$.

$$\mu = \frac{1}{N}\sum_{k=1}^{N} X_k = \frac{2+5+8+13}{4} = 7$$

Population standard deviation:

$$\sigma = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(X_k-\mu)^2} = \sqrt{\frac{(2-7)^2+(5-7)^2+(8-7)^2+(13-7)^2}{N}}$$

$$= \sqrt{\frac{25+4+1+36}{4}} = \sqrt{\frac{66}{4}} = \sqrt{16.5} = 4.06$$

ii) (a) Number of possible samples with replacement = $N^n = 16$.

The samples:

(2,2), (2,5), (2,8), (2,13),
(5,2), (5,5), (5,8), (5,13),
(8,2), (8,5), (8,8), (8,13),
(13,2), (13,5), (13,8), (13,13).

The sample means:

2, 3.5, 5, 7.5,
3.5, 5, 6.5, 9,
5, 6.5, 8, 10.5,
7.5, 9, 10.5, 13.

Sampling Distribution of Sample Means:

$\bar{x}$	f	$\frac{f}{N}$ =P(Probability)
2	1	$\frac{1}{16}$
3.5	2	$\frac{2}{16}$
5	3	$\frac{3}{16}$
6.5	2	$\frac{2}{16}$
7.5	2	$\frac{2}{16}$
8	1	$\frac{1}{16}$
9	2	$\frac{2}{16}$
10.5	2	$\frac{2}{16}$
13	1	$\frac{1}{16}$

Mean of the sampling distribution:

$$E(\bar{x}) = \sum_{i=1}^{n} P_i\bar{x}_i \text{ (where } \bar{x}_i \text{ is the mean of } i^{th} \text{ sample)}$$

$$= \frac{1}{16}(2+7+15+13+15+8+18+21+13)$$

$$= \frac{1}{16} \times 112 = 7$$

Standard error of the distribution:

$$\sigma_{\bar{x}} = \sqrt{E(\bar{x}^2) - \{E(\bar{x})\}^2}$$

Now,

$$E(\bar{x}^2) = \sum_{i=1}^{n} P_i\bar{x}_i^2$$

$$= \frac{1}{16}(1\times2^2 + 2\times3.5^2 + 3\times5^2 + 2\times6.5^2 + 2\times7.5^2 + 1\times8^2 + 2\times9^2 + 2\times10.5^2 + 1\times13^2)$$

$$= \frac{1}{16}(1 \times 4 + 2 \times 12.5 + 3 \times 25 + 2 \times 42.25 + 2 \times 56.25 + 1 \times 64 + 2 \times 81 + 2 \times 110.5 + 1 \times 169)$$

$$= \frac{1}{16}(4+25+75+84.5+112.5+64+81+221+169)$$

$$= \frac{1}{16} \times 748 = 57.31.$$

And,

$$\{E(\bar{x})\}^2 = 7^2 = 49$$

$$\sigma_{\bar{x}} = \sqrt{57.31 - 49} = \sqrt{8.31} = 2.83.$$

Thus, the mean and the standard error of the sampling distribution in the case of random sampling with replacement are 7 and 2.83 respectively.

b) Number of possible samples without replacement = ${}^{N}C_n = {}^{4}C_2 = 6$

The samples:

(2,5), (2,8), (2,13), (5,8), (5,13), (8,13).

Sample means:

3.5, 5, 7.5, 6.5, 9, 10.5.

Sampling Distribution of Sample Means:

$\bar{x}$	f	$\frac{f}{N}$ =P(Probability)
3.5	1	$\frac{1}{6}$
5	1	$\frac{1}{6}$

7.5	1	$\frac{1}{6}$
6.5	1	$\frac{1}{6}$
9	1	$\frac{1}{6}$
10.5	1	$\frac{1}{6}$

Mean of the population:

$$E(\bar{x}) = \sum_{i=1}^{n} P_i \bar{x}_i$$

$$= \frac{1}{6}(3.5+5+7.5+6.5+9+10.5)$$

$$= \frac{1}{6}\times 42 = 7.$$

Standard error of the distribution

$$\sigma_{\bar{x}} = \sqrt{E(\bar{x}^2) - \{E(\bar{x})\}^2}$$

Now,

$$E(\bar{x}^2) = \sum_{i=1}^{n} P_i x_i^2$$

$$= \frac{1}{6}(3.5^2 + 5^2 + 7.5^2 + 6.5^2 + 9^2 + 10.5^2)$$

$$= \frac{1}{6}(12.25 + 25 + 56.25 + 42.25 + 81 + 110.25)$$

$$= \frac{1}{6}\times 327 = 54.5$$

And we already know,

$$\{E(\bar{x})\}^2 = 7^2 = 49$$

$$\sigma_{\bar{x}} = \sqrt{54.5 = 49} = \sqrt{5.5} = 2.35$$

Thus, the mean and the standard error of the sampling distribution in the case of random sampling without replacement are 7 and 2.35 respectively.

iii) In the case of random sampling with replacement,

$$E(\bar{x}) = 7 = \mu$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{4.06}{\sqrt{2}} = \frac{4.06}{1.414} = 2.87 \cong 2.83,$$

as we have independently obtained from the sampling distribution of the sample means.

In the case of random sampling without replacement,

$E(\bar{x}) = 7 = \mu$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \times \sqrt{\frac{N-n}{N-1}} = \sqrt{\frac{\sigma^2}{n} \times \frac{N-n}{N-1}} = \sqrt{\frac{16.48}{2} \times \frac{2}{3}} = \sqrt{8.24 \times 0.67} = \sqrt{5.52} = 2.35,$$

as we have independently obtained from the sampling distribution of the sample mean

Hence, our results are all verified.

16.9 DESIRABLE PROPERTIES OF AN ESTIMATOR

Suppose, θ is an unknown population *parameter* that we are interested in. We may want to estimate θ on the basis of a random sample drawn from the population. For this purpose we may use a statistic T (which is a function of the sample values). Here T is an *estimator* of θ and the value of T that is obtained from the given sample is an *estimate* of θ. In fact, the value is known as a *point estimate* in the sense that it is one particular value of the estimator (see Unit 18 for details).

Earlier, we have discussed the concepts of sampling and non-sampling errors. We recapitulate here that the absolute difference (ignoring the sign) between a sample statistic and the population parameter, i.e., $|T-\theta|$ measures the extent of the sampling error. We may note here that an estimator is essentially a formula for computing an estimate of the population parameter and there can be several potential estimators (alternative formulae) that may be used for this purpose. So, there should be some desirable properties on the basis of which we can select a particular estimator for estimating the population parameter. A very simple requirement for T to be a good estimator of θ is that the difference $|T-\theta|$ should be as small as possible. Various approaches have been suggested to ensure this.

16.9.1 Unbiasedness

We have already noted that the value of a statistic varies from sample to sample due to sampling fluctuation. Although the individual values of a statistic may be different from the unknown population parameter, on an average, the value of a statistic should be equal to the population parameter. In other words, the sampling distribution of T should have a central tendency towards θ. This is known as the property of unbiasedness of an estimator. It means that although an individual value of a given estimator may be higher or lower than the unknown value of the population parameter, there is no bias on the part of the estimator to have values that are always greater or smaller than the unknown population parameter. If we accept that mean (here, expectation) is a proper measure for central tendency, then T is an *unbiased estimator* for θ if $E(T) = \theta$.

16.9.2 Minimum-Variance

It is also desirable that the average spread of all the possible values of an unbiased estimator around the population parameter is as small as possible. It will reduce the chance of an estimate being far away from the parameter. If we accept that variance is a proper measure for average spread (dispersion), we want that among all the unbiased estimators, T should have the smallest variance. Symbolically, $V(T) \leq V(T')$ where, V stands for variance and T' is any other unbiased estimator.

An estimator T, which is unbiased and among all the unbiased estimator has the minimum variance, is known as a *minimum-variance unbiased estimator.* Let us consider an example. Suppose, we have a random sample of size n from a given population of size N. In this case, the sample mean is given by $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ where x_i is the i^{th} member of the sample. It can be proved that it is an unbiased estimator of the population mean μ. Symbolically

$E(\bar{x}) = \mu$

However, it can be shown that the sample variance defined as $s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ is not an unbiased estimator of the population variance σ^2. Symbolically,

$E(s^2) \neq \sigma^2$

On the contrary, if we define the sample variance as $s'^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$, then s'^2 is an unbiased estimator of $\sigma'^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2$.

Suppose further that the sample values are not only random but also independent (random sample with replacement) and the underlying population is normal. It can be shown that the sample mean $\bar{x}$ is not only an unbiased estimator of the population mean μ but also it has the minimum variance among all the unbiased estimators of μ.

16.9.3 Consistency and Efficiency

Another approach may be to suggest that the estimator T should approximate the unknown population parameter θ as the sample size n increases. Since T itself is a random variable, we may express this requirement in probabilistic or stochastic terms as the statistic T should converge to the parameter θ stochastically (i.e., in probability) as $n \to \infty$. A statistic T with this property is called a *consistent* estimator of θ.

In real life, a large number of consistent estimators of the same parameter q have often been found. In such a situation, obviously, some additional criterion is needed to choose among these consistent estimators. One such criterion may be to demand that not only T should converge stochastically to θ but also it should do so quite rapidly. Without going into the details, we may mention here that some times an estimator assumes the form of a normal distribution when the sample size n increases indefinitely. Such estimators are called *asymptotically normal.* If we focus on consistent estimators that are asymptotically normal, the rapidity of their convergence is indicated by their respective asymptotic variances. In fact, the convergence is the fastest for the estimator that has the *lowest asymptotic variance.* Such kind of an estimator is known as an *efficient estimator* among all the asymptotically normal consistent estimators of a population parameter.

Check Your Progress 2

1) Define the following concepts:

   a) Simple Random Sample

b) Sampling Distribution

c) Standard Error

....

....

....

....

2) Distinguish between the following:

a) Simple random sampling with replacement and Simple random sampling without replacement

b) Simple random sampling and stratified random sampling

....

....

....

....

3) Given a population: 1, 2, 5, 6. Bring out all possible samples of size 2

i) with replacement, and

ii) without replacement.

....

....

....

....

4) Given a population: 2, 4, 6. Suppose a sample of size 2 is to be selected from this population by the method of random sampling without replacement.

a) Present the sampling distribution of sample mean.

b) Compute the standard error.

....

....

....

....

16.10 LET US SUM UP

In this unit, we distinguished between the census method and the sample method of conducting a statistical inquiry. We have seen that on account of various resource constraints, census method cannot be undertaken always. Moreover, due to the enormity of the task involved, the chances of committing non-sampling errors in a census are at times quite high. A sample survey, on the other hand, has some definite advantages. A properly conducted sample survey is generally less error prone. A sample survey has a sound scientific basis. As a result, the sampling distribution of the relevant statistic (obtained from random samples) forms an objective basis of assessment about a population parameter.

16.11 KEY WORDS

Estimate : It is the particular value that can be obtained from an estimator.

Estimator : It is the specific functional form of a statistic or the formula involved in its calculation. Generally, the two terms, statistic and estimator, are used interchangeably.

Parameter : It is a measure of some characteristic of the population.

Population : It is the entire collection of units of a specified type in a given place and at a particular point of time.

Random Sampling : It is a procedure where every member of the population has a definite chance or probability of being selected in the sample. It is also called probability sampling.

Sample : It is a sub-set of the population. Therefore, it is a collection of some units from the population.

Sampling Distribution : It refers to the probability distribution of a statistic.

Sampling Error : The absolute difference between population parameter and relevant sample statistic.

Sampling Fluctuation : It is the variation in the values of a statistic computed from different samples.

Simple Random Sampling : This is a sampling procedure, in which, each member of the population has the *same chance* of being selected in the sample.

Standard Error : It is the standard deviation of the sampling distribution of a statistic.

Statistic : It is a function of the values of the units that are included in the sample. The basic purpose of a statistic is to estimate some population parameter.

Statistical Inference : It is the process of drawing conclusions about an unknown population characteristic on the basis of a known sample drawn from it.

16.12 SOME USEFUL BOOKS

Bhardwaj. R. S., 1999, *Business Statistics* (First Edition), Excel Books, New Delhi, Chapter 20.

Nagar, A. L. and Das, R. K., 1988, *Basic Statistics,* Oxford University Press, Delhi, Chapter 9.

Goon, A. M., Gupta, M. K. and Dasgupta, B., 1971, *Fundamentals of Statistics,* Volume 1 (Fourth Edition), The World Press Private Limited, Calcutta, Chapters 14, 15 and 16.

16.13 ANSWERS/HINTS TO CHECK YOUR PROGRESS EXERCISES

Check Your Progress 1

1) Read the text and define these terms in one or two sentences each.
2) Read the text and distinguish in a few sentences.
3) Read the text and answer in a few sentences.

Check Your Progress 2

1) Read the text and define these terms in a few sentences.
2) Read the text and distinguish in a few sentences.
3) Go through Example 16.2 in the text and attempt yourself.
4) 0.82.

Chi-square Distribution

A square of a standard normal variate is known as a chi-square variate with 1 degree of freedom. If $X \sim N(\mu, \sigma^2)$ Then $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$ and $Z^2 = \left(\frac{X-\mu}{\sigma}\right)^2$ is a chi-square variate with 1 degree of freedom (d. f).

If x_i (i = 1,2…n) are n independent normal variates with means μ_i and σ_i^2 variances then $\chi^2 = \sum_{i=1}^{n}\left(\frac{x_i-\mu_i}{\sigma_i}\right)^2$ is a chi-square variate with 'n' degree of freedom..

Chi-square Distribution

The Probability density function of chi-square distribution with 'n' degrees of freedom is given by

$$f(x) = \begin{cases} \dfrac{1}{2^{\frac{n}{2}}\Gamma\left(\dfrac{n}{2}\right)} e^{-\frac{x}{2}} x^{\frac{n}{2}-1} , & 0 < x < \infty \\ 0 , & other\ wise \end{cases}$$

Remark: If $X \sim \chi_n^2$ **then** $\frac{1}{2}X \sim \gamma\left(\frac{n}{2}\right)$

Chi-square Distribution

Properties:

1. If n = 1 normal distribution is a particular case of Chi-Square distribution.
2. If n = 2 exponential distribution is a particular case of Chi – Square distribution.
3. Mean = n , Variance = 2n, Mode = (n-2)
4. $M_x(t) = (1-2t)^{-\frac{n}{2}}, |2t| < 1.$
5. Measures of skewness: $\beta_1 = \frac{8}{n} \Rightarrow \gamma_1 = 2\sqrt{\frac{2}{n}}$
6. Measures of kurtosis: $\beta_2 = 3 + 12/n \Rightarrow \gamma_2 = 12/n$

Chi-square Distribution

Applications:

1. To test if the hypothetical value of the population variance is
2. To test the goodness of fit.
3. To test the independence of attributes.
4. To test the homogeneity of independent estimates of the population variance.
5. To test the equality of several population correlation coefficients.

‘t’-Distribution – Student’s ‘t’ distribution

If the sample size n is small, the distribution of the various statistics $Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$ or $Z = \frac{X - np}{\sqrt{nPQ}}$ are far from normality

and as such ‘normal test’ cannot applied if n is small. In such cases exact sample tests, pioneered by W.S.Gosset (1908) who wrote under the pen name of Student, and later on developed and extended by Prof. R. A .Fisher(1926),are used.

Assumptions :

1. The parent population from which the sample is drawn is normal
2. The population observations are Independent.
3. The sample variance s^2 is unknown.

't'-Distribution – Student's 't' distribution

Let $X_i (i = 1, 2, ..., n)$ be a random sample of size n from a normal population with mean μ and σ^2 variance . Then student t is defined by the statistic $t = \frac{\bar{x} - \mu}{S / \sqrt{n}}$ where $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ is the sample mean and S^2 is an unbiased estimator of the population variance σ^2 and it follows student's 't' – distribution with (n -1) degrees of freedom.

The Probability density function of Students' t- distribution is given by

$$f(t) = \begin{cases} \dfrac{1}{\sqrt{\nu}\,\beta\left(\dfrac{1}{2}, \dfrac{\nu}{2}\right)} \cdot \dfrac{1}{\left(1 + \dfrac{t^2}{\nu}\right)^{\frac{(\nu+1)}{2}}}, -\infty < t < \infty \\ 0 \qquad other\ \ wise; \nu = n - 1 \end{cases}$$

‘t’-Distribution – Student’s ‘t’ distribution

Properties

1. Students ‘t’ is a particular case of fisher’s ‘t’.
2. ‘t’ distribution is symmetrical about the origin.
3. All moments of odd order are zero
4. The mean of t distribution is zero
5. Variance = $\mu_2 = \frac{n}{n-2}$
6. Measure of skewness = $\beta_1 = 0$
7. Measure of kurtosis = $\beta_2 = 3\left(\frac{n-2}{n-4}\right)\ ;\ n > 4$
8. Mode of t – distribution lies at the origin t = 0
9. M.G.F of ‘t’ -distribution does not exist.

t'-Distribution – Student's 't' distribution

Application

1. t – distribution is used to test whether the sample mean to a specified value of population mean or not.
2. To test the significance of the difference between two sample means.
3. To test the significance of an observed sample correlation coefficients and sample regression coefficient.
4. To test the significance of observed partial correlation coefficient.

F – Distribution (Snedecor's F distribution)

If X and Y are two independent chi-square variates with & degrees of freedom $\nu_1 \& \nu_2$ respectively, then F – statistic is defined by $F = \frac{X/\nu_1}{Y/\nu_2}$ i.e. F is defined as the ratio of two independent chi-square variates divided by their respective degrees of freedom and it follows Snedecor's F distribution with (ν_1, ν_2) degrees of freedom with probability density function is given by

$$f(x) = \frac{\left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}}}{\beta\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right)} \cdot \frac{x^{\frac{\nu_1}{2}-1}}{\left(1 + \frac{\nu_1}{\nu_2}x\right)^{(\nu_1+\nu_2)/2}}, 0 \leq x < \infty$$

$$\nu_1 = n_1 - 1 \,\&\, \nu_2 = n_2 - 1$$

F – Distribution (Snedecor's F distribution)

Properties

1. F – distribution extends along abscissa from 0 to .
2. F – distribution curve wholly lies in the first quadrant.
3. Shape of the curve depends on the degrees of freedom.
4. F – distribution curve is a positive skew curve and is highly positively skewed when is small $\nu_2 < 5$
5. The curve is unimodel and its mode is at the point $X = \frac{\nu_2(\nu_1 - 2)}{\nu_1(\nu_2 + 2)}$ mode is always less than unity for $\nu_2 > 2$
6. Mean of F – distribution is $\frac{\nu_2}{\nu_2 - 2}$
7. Variance of F – distribution is $\frac{2\nu_2^2(\nu_2 + \nu_1 - 2)}{\nu_1(\nu_2 - 2)^2(\nu_2 - 4)}, \ \nu_2 > 4$
8. M.g.f. of F – distribution does not exist.
9. F distribution is highly positively skewed.

F – Distribution (Snedecor's F distribution)

Applications

1. F-distribution is used to test the equality of two population variances.
2. F-distribution is used to analysis of variance.
3. To test the equality of several regression coefficients.
4. To test the equality of several population means.
5. To test the significance of an observed multiple correlation coefficient.
6. To test the significance of an observed sample correlation ratio.

Relation between t, F and χ^2

If t has d.f and F has $(1, n)$ degrees of freedom then $t_n^2 = F_{1,\,n}$

In $F(n_1, n_2)$ distribution if we let $n_2 \to \infty$ then

$$\boxed{\chi^2 = n_1 F \sim \chi^2_{n_1}}$$

TEST OF HYPOTHESIS

POPULATION: The collection of objects about a particular characteristic which we are going to study is called population

Example: 1. All engineering colleges Andhra Pradesh.

2. Number of vehicles in Tirupati.

POPULATION SIZE: Number of objects in the given population is defined as population size. It is denoted with 'N'**.** Depend upon population size, There are two types

1. Finite population
2. Infinite population

1. Finite population:

A population is said to be finite if the population size is countable

Example: Number of pens

2. Infinite population

A population is said to be infinite if the population size is uncountable.

Example: Stars in the sky

SAMPLE: A finite subset of a population is called a sample

Example: 1. Number of engineering colleges in Tirupati.

2. Number of two-wheelers in Tirupati.

SAMPLE SIZE: Number of objects in a given sample is defined as sample size. It is denoted with 'n'. Depend upon Sample size, there are two types

1. Large sample
2. Small sample

1. Large Sample:

If the sample size $n \geq 30$, then the sample is called a large sample

2. Small Sample:

If the sample size $n < 30$, then the sample is called a small sample

PARAMETER: The statistical constants obtained by using population data is called a parameter

Example: 1. Population mean = μ
2. Population variance = σ^2
3. Population Standard Deviation = σ

STATISTIC: The statistical constants obtained by using statistical data is called statistic

Example: 1. Sample mean = $\overline{X}$ (or) μ_0
2. Sample variance = S^2
3. Sample Standard Deviation = S

STATISTICAL HYPOTHESIS :

- To make decisions about a population with the basis of sample information. Such decisions may or may not be true are called a statistical hypothesis
- It is simply a statement about the statistical hypothesis. They are of two types
  1. Null hypothesis
  2. Alternative hypothesis

1. **Null Hypothesis:** A definite statement about the population parameter is called the null hypothesis (H_0). H_0 is a commonly accepted fact.

A null hypothesis the hypothesis which asserts that there is no significant difference between the statistic and the population parameter and whatever observed difference is there is merely due to fluctuations in sampling from the same population.

$$H_0 \rightarrow \mu = \mu_0 \ ;\sigma = \sigma_0$$

2. **Alternative hypothesis**: A hypothesis which is complementary to the null hypothesis is called the alternative hypothesis(H_1).

$$H_1 \rightarrow \mu \neq \mu_0 \ ;\sigma \neq \sigma_0$$

$$\rightarrow \mu > \mu_0$$

$$\rightarrow \mu < \mu_{0\,:}$$

- Based on the alternative hypothesis, the test is divided into two types
  i. Two-tailed test
  ii. One-tailed test

- **Two-Tailed Test:**

If the alternative hypothesis is “not equal to” type, then the critical region completely lies on both sides of the curve having area ‘α/2’ on both sides.

- **One-tailedtest:**One-tailed test is of two types

a) Right-tailed test

b) Left-tailed test

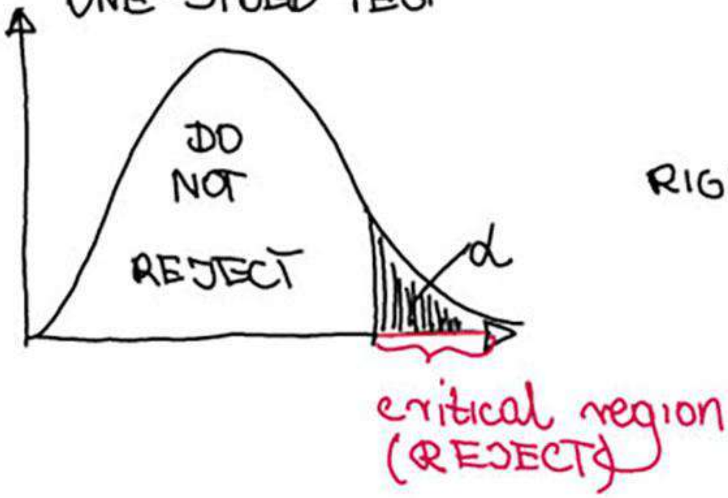
- **Right-Tailed Test:**

If the alternative hypothesis is of “greater than“ type($H_1 : \mu > \mu_0$), then the critical region completely lies on the right side of the curve having area ‘α ’.

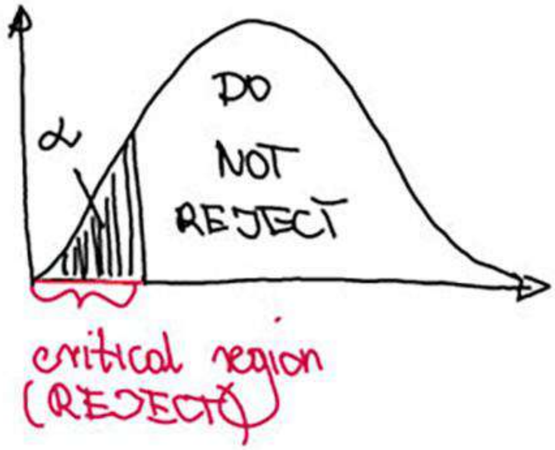
- **Left-tailed test:**

If the alternative hypothesis is of “less than” type ($H_1 : \mu < \mu_0$), then the critical region completely lies on the left side of the curve-handling area ‘α ’.

ONE-SIDED TEST

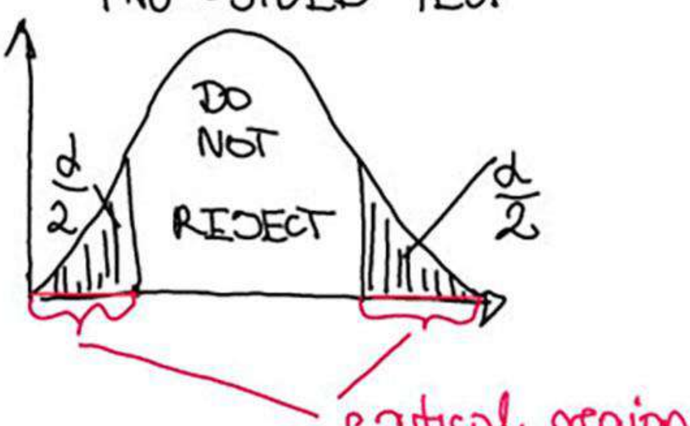


RIGHT



LEFT

TWO-SIDED TEST



ERRORS IN SAMPLING:

The main objective of sampling theory is to draw or obtain a valid inference (information) about population When a statistical hypothesis is tested, there are four possibilities.

1. The Hypothesis is true and the Test is accepted (Correct Decision)
2. The Hypothesis is false and the Test is rejected (Correct Decision)
3. The Hypothesis is true and the Test is rejected (Type1 Error)
4. The Hypothesis is false and the Test is accepted (Type 2 Error)

This can be shown in the following table

H_0	Accept	Reject
True	Correct Decision	Type 1 Error (Producer's Risk)
False	Type 2 Error (Consumer's Risk)	Correct Decision

Note:

1. Type 1 Error (Probability 'α'): The error involves rejecting H_0 even though it is true. Is also known as "Producer's Risk".

2. Type 2 Error (Probability 'β'): This error involvesaccepting H_0 even though it is false. This error is also known as "Consumer's Risk".

- **CRITICAL REGION:**

  The region in which the null hypothesis is rejected is called critical region.

- **ACCEPTANCE REGION**

  The region in which the null hypothesis is accepted is called the acceptance region.

- **LEVEL OF SIGNIFICANCE**

  The maximum probability of Type 1 error isα. Here 'α' is called Level of Significance.

  - It generally expressed in '%'.
  - It measures errors associated with making decisions.

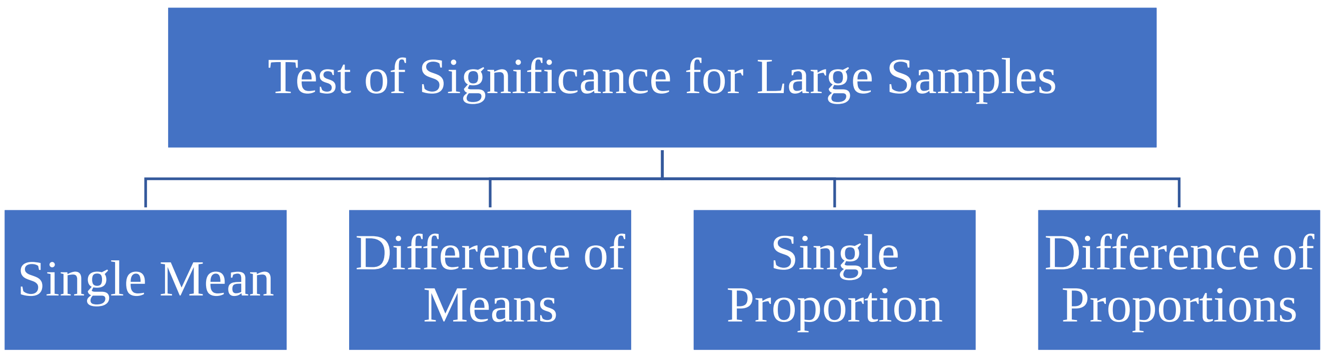
  We take α=1%, which is used for high accuracy.

  α =5%, which is used for moderate accuracy

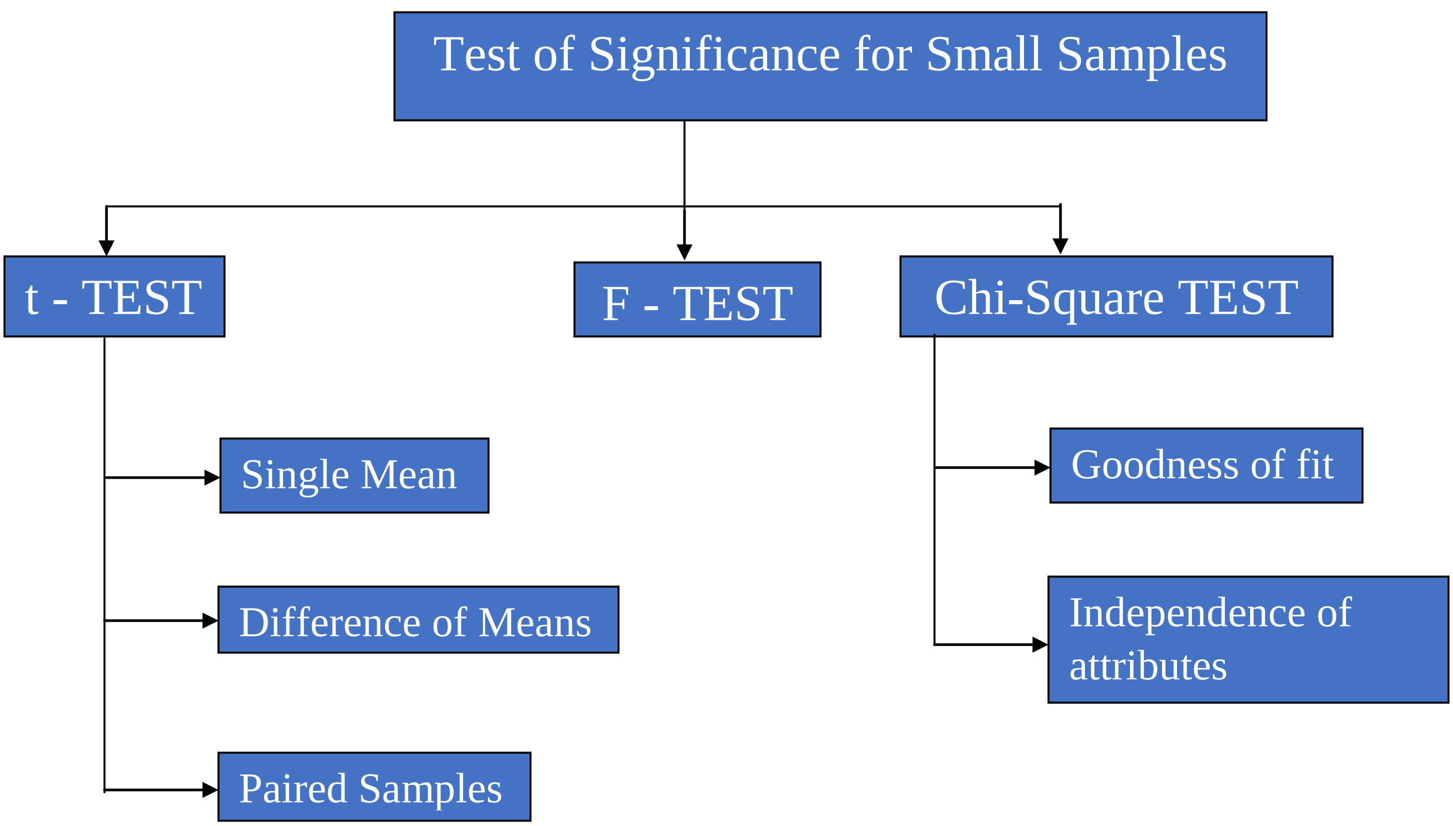
Procedure for Testing of Hypothesis

1. Setup Null Hypothesis
2. Setup Alternate Hypothesis
3. Fix Level of Significance
4. Choose suitable Test Statistic
5. Interpretation

UNIT-II



UNIT-III



UNIT-II

Z Test for Single Mean

Let $\bar{x}$ be the sample mean of the large sample size *'n'* drawn from the normal population having mean μ and standard deviation 'σ'

Null Hypothesis H_0:

There is no significant difference between the sample mean and population mean $\bar{x} = \mu$ (or)

The sample is drawn from the normal population

Alternative Hypothesis H_1:

There is a significant difference between the sample mean and population mean

$$\bar{x} \neq \mu \, (TwoTailed)$$

$$\left.\begin{array}{l} \bar{x} < \mu \\ \bar{x} > \mu \end{array}\right\} (OneTailed) \quad \text{(or)}$$

The sample is not drawn from the normal population

Level of Significance α:

Choose the level of significance based on the problem. Default 5%

Test Statistic:

SE=Standard Error $\quad z = \dfrac{\left|\bar{x} - \mu\right|}{SE\left(\bar{x}\right)}; \quad z = \dfrac{\left|\bar{x} - \mu\right|}{\sigma / \sqrt{n}}, \, if \; \sigma \, is \, known; \quad z = \dfrac{\left|\bar{x} - \mu\right|}{S / \sqrt{n}}, \, if \; \sigma \, is \, not \, known$

Conclusion:

If Z-calculated value < Z critical value or table value at α% level of significance then we accept H_0.

Otherwise we reject H_0

Confidence Limits or Fudicial limits: $\quad Z = \bar{x} \pm Z_\alpha SE\left(\bar{x}\right)$

Problems:

1. A sample of 64 students have a mean weight of 70kgs, can this be regarded as a sample from a population with mean weight 56 kgs and standard deviation 25kgs.
2. An oceanographer wants to check whether the depth of the ocean in a certain region is 57.4 fathoms as had previously been recorded. What can he conclude at 0.05 level of significance, if readings taken at 40 random locations in the given region yielded a mean of 59.1 fathoms with a standard deviation of 5.2 fathoms.
3. A sample of 900 members has a mean of 3.4 cms and S.D. 2.61 cms. Is this sample has been taken from a large population of mean 3.25 cms and S.D.2.61 cms. If the population is normal then test the hypothesis with 95% fudicial limits.

4. A sample of 400 items is taken from a population whose S.D. is 10. The mean of the sample is 40, test whether the sample has come from a population with mean 38. Also calculate 95% confidence limits for the population.
5. An ambulance service claims that it takes on the average less than 10 minutes to reach its destination in emergency calls. A sample of 36 calls has a mean of 11 minutes and the variance of 16 minutes. Test the claim at 0.05 level of significance.
6. An insurance agent has claimed that the average age of policy holders who issue through him is less than the average for all agents which is 30.5 years. A random sample of 100 policy holders who had issues through him gave the following age distribution.

Age	16-20	21-25	26-30	31-35	36-40
No. of Persons	12	22	20	30	16

Z Test for Difference of two Means

Let $\bar{x}_1$ *and* $\bar{x}_2$ be the two sample means of the large sample sizes *'n_1'* and *'n_2'* drawn from two normal populations having mean μ_1 *and* μ_2 and standard deviation σ_1 *and* σ_2

Null Hypothesis H_0:

There is no significant difference between the two population means $\mu_1 = \mu_2$ (or)

The two samples are drawn from same population

Alternative Hypothesis H_1:

There is a significant difference between the two population means

$$\mu_1 \neq \mu_2 \; (TwoTailed)$$
$$\left.\begin{array}{l}\mu_1 < \mu_2 \\ \mu_1 > \mu_2\end{array}\right\} (OneTailed) \quad \text{(or)}$$

The two samples are not drawn from same population

Level of Significance α:

Choose the level of significance based on the problem. Default 5%

Test Statistic: SE=Standard Error

$$z = \frac{\left|\bar{x}_1 - \bar{x}_2\right|}{SE\left(\bar{x}_1 - \bar{x}_2\right)};$$

$$z = \frac{\left|\bar{x}_1 - \bar{x}_2\right|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}, \; if \; \sigma_1, \sigma_2 \; is \, known \; ; \quad z = \frac{\left|\bar{x}_1 - \bar{x}_2\right|}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}, \; if \; \sigma_1, \sigma_2 \; is \, not \, known;$$

$$z = \frac{\left|\bar{x}_1 - \bar{x}_2\right|}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \; if \; \sigma_1 = \sigma_2 = \sigma \; is \, known$$

Conclusion:

If Z-calculated value < Z critical value or table value at α% level of significance then we accept H_0.

Otherwise we reject H_0

Confidence Limits or Fudicial limits: $Z = \left|\bar{x}_1 - \bar{x}_2\right| \pm Z_\alpha SE\left(\bar{x}_1 - \bar{x}_2\right)$

Problems:

1. The mean yield of wheat from a district A was 210 pounds with S.D. 10 pounds per acre from a sample of 100 plots. In another district B the mean yield was 220 pounds with S.D. 12 pounds from a sample of 150 plots. Assuming that the S.D. of yield in the entire state was 11 pounds, test whether there is any significant difference between the mean yields of crops in the two districts.
2. The research investigator is interested in studying whether there is a significant difference in the salaries of MBA grades in two metropolitan cities. A random sample of size 100 from Mumbai yields on average income of Rs. 20,150/-. Another random sample of 60 from Chennai results in an average income of Rs. 20,250/-. If the variances of both the populations are given as Rs. 40,000/- and Rs. 32,400/- respectively.
3. A simple sample of the height of 6400 Englishmen has a mean of 67.85 inches and a S.D. of 2.56 inches while a simple sample of heights of 1600 Austrians has a mean of 68.55 inches and S.D. of 2.52 inches. Do the data indicate the Austrians on the average taller than the Englishmen? (Use 0.01).
4. The average mark scored by 32 boys is 72 with a SD of 8, while that for 36 girls is 70 with a SD of 6. Does this indicate that the boys perform better than girls at level of significance 0.05?
5. A company claims that its bulbs are superior to those of its main competitor. If a study showed that a sample 40 of its bulbs have a mean life time of 647 hrs of continuous use with a SD of 27 hrs. While a sample of 40 bulbs made by its main competitor had a mean life time of 638 hrs of continuous use with a SD of 31 hrs. Test the significance between the differences of two means at 5% level and fudicial limits.

Z Test for Single Proportion

Let *'p'* be the sample proportion of the large sample size *'n'* drawn from the normal population having proportion *'P'*

Null Hypothesis H_0:

There is no significant difference between the sample proportion and population proportion $p = P$ (or)

The sample is drawn from the normal population

Alternative Hypothesis H_1:

There is a significant difference between the sample proportion and population proportion

$p \neq P$ (Two Tailed);

$p < P$ (or) $p > P$ (One Tailed)

(or)

The sample is not drawn from the normal population

Level of Significance α:

Choose the level of significance based on the problem. Default 5%

Test Statistic:

SE=Standard Error   Q = 1–P, q = 1–p

$$z=\frac{|p-P|}{SE(p)};\quad z=\frac{|p-P|}{\sqrt{PQ/n}},\ if\ P\ is\ known\ ;\quad z=\frac{|p-P|}{\sqrt{pq/n}},\ if\ P\ is\ not\ known;$$

Conclusion:

If Z-calculated value < Z critical value or table value at α% level of significance then we accept H_0.

Otherwise we reject H_0

Confidence Limits or Fudicial limits: $Z = p \pm Z_{\alpha} SE(p)$

Problems:

1. A manufacturer claimed that atleast 95% of the equipment which he supplied to a factory confirmed to specifications. An examinations of sample of 200 pieces of equipment revealed that 18 were faulty. Test his claim at 5% level of significance.
   Given n=200 ; P=95%=0.95; p=182/200=0.91; Q=5%=0.05
   P=5%=0.05; p=18/200=0.09, Q=0.95
   Ho: mu=95%
   H1: mu>95%
   Alpha= 5%
   Z=2.5974
   Z table value=1.645
   Z cal > z table then reject Ho

2. In a sample of 1000 people in Karnataka 540 are rice eaters and the rest are wheat eaters. Can we assume that both rice and wheat are equally popular in this state at 1% level of significance?
   n=1000; P= 0.5; p= 540/1000=0.54; Q=0.5
   Z=2.5316
   Z table value 2.58

3. In a big city 325 men out of 600 men were found to be smokers. Does this information support the conclusion that the majority of men in this city are smokers?
4. A die was thrown 9000 times and of these 3220 yielded as 3 or 4. Is this consistent with the hypothesis that the die was unbiased?
   n=9000   U + or   n*and

   p= 3220/9000

   P=1/6+1/6=1/3=0.3333
5. A random sample of 500 pineapples was taken from a large consignment and 65 were found to be bad. Find the percentage of bad pineapples in the consignment.
6. A social worker believes that fewer than 25% of the couples in a certain area have ever used any form of birth control. A random sample of 120 couples was contacted, twenty of them said that they have used. Test the belief of the social worker at 0.05 level?

n=120; p=20/120; P=25%; Q=75%

Z Test for Difference of two Proportions

Let p_1 and p_2 be the two sample proportions of the large sample sizes *'n_1'*and *'n_2'* drawn from two normal populations having proportions P_1 and P_2

Null Hypothesis H_0:

There is no significant difference between the two population proportions $P_1 = P_2$ (or)

The two samples are drawn from same population

Alternative Hypothesis H_1:

There is a significant difference between the two population proportions

$P_1 \neq P_2$ (Two tailed) or

$P_1 < P_2$ or $P_1 > P_2$ (One tailed)

(or)

The two samples are not drawn from same population

Level of Significance α:

Choose the level of significance based on the problem. Default 5%

Test Statistic: SE=Standard Error $Q_1 = 1–P_1$, $q_1 = 1–p_1$, $Q_2 = 1–P_2$, $q_2 = 1–p_2$

$$z = \frac{|p_1 - p_2|}{SE(p_1 - p_2)};$$

$$z = \frac{|p_1 - p_2|}{\sqrt{\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}}}, \; if \; P_1, P_2 \; is \, known \;; \; z = \frac{|p_1 - p_2|}{\sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}}, \; if \; P_1, P_2 \; is \, not \, known;$$

$$z = \frac{|p_1 - p_2|}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}, \; if \; P_1, P_2 \; is \, not \, known \; (Method \; of \; pooling)$$

$$where, \; p = \frac{n_1p_1 + n_2p_2}{n_1 + n_2}$$

Conclusion:

If Z-calculated value < Z critical value or table value at α% level of significance then we accept H_0.

Otherwise we reject H_0

Confidence Limits or Fudicial limits: $Z = |p_1 - p_2| \pm Z_\alpha SE(p_1 - p_2)$

Problems:

1. Random samples of 400 men and 600 women were asked whether they would like to have a flyover neat their residence. 200 men and 325 women were in favour of the proposal. Test the hypothesis that proportions of men and women in favour of the proposal are same at 5% level.
2. On the basis of their total scores, 200 candidates of a civil service examination are divided into two groups, the upper 30% and the remaining 70%. Consider the first question of the examination. Among the first group 40 had the correct answer whereas among the second group 80 had the correct answer. On the basis of these results can one conclude that the first question is not good at discriminating ability of the type being examined here?
3. In a city A, 20% of a random sample of 900 school boys has a certain slight physical defect. In another city B, 18.5% of a random sample of 1600 school boys has the same defect. Is the difference between the proportions significant at 0.05 level of significance.
4. A machine puts out 9 imperfect articles in a sample of 200 articles. After the machine is overhauled (service) it puts out 5 imperfect articles in a sample of 700 articles. Test at 5% level whether the machine is improved?
5. Before an increase on excise duty on tea 500 people out of a sample of 900 found to have the habit of having tea. After an increase on excise duty 250 are have the habit of having tea among 1100. Is there any decrease in the consumption of tea. Test at 5% level.
6. During a country wide investigation the incidence of tuberculosis was found be 1%. In a college of 400 students 3 reported to be affected whereas in another college of 1200 students 10 were affected. Does this indicate any significant difference?

Unit-III

Degrees of Freedom:

It is a number which indicates how many of the values of a variable may be independently or freely chosen.

Example:

If we have chosen any four numbers freely then we may choose 11, 6, 14, 28 or any other set of four numbers. As in this case all the four have freedom to vary we say that the degree of freedom is 4.

If we impose a restriction on the numbers say the sum is 50 then we can choose three numbers freely and the fourth number is such that the sum is 50. As in this case three numbers are free and fourth number is dependent then the degrees of freedom is 3 (4–1)

In general, the number of degrees of freedom is equal to the total number of observations less the number of independent constraints imposed on the observations.

Student's t – test for Single Mean

Let $\bar{x}$ be the sample mean of the small sample size *'n'* drawn from the normal population having mean μ and standard deviation 'σ'

Null Hypothesis H_0:

There is no significant difference between the sample mean and population mean $\bar{x} = \mu$ (or)

The sample is drawn from the normal population

Alternative Hypothesis H_1:

There is a significant difference between the sample mean and population mean

$$\bar{x} \neq \mu \,(TwoTailed)$$
$$\left.\begin{array}{l}\bar{x} < \mu \\ \bar{x} > \mu\end{array}\right\}(OneTailed) \quad \text{(or)}$$

The sample is not drawn from the normal population

Level of Significance α:

Choose the level of significance based on the problem. Default 5%

Test Statistic:

$$t = \frac{\left|\bar{x} - \mu\right|}{SE\left(\bar{x}\right)}; \ \sim t - distribution\ with\ \ n-1 \deg rees\ of\ \ freedom$$

SE=Standard Error $\quad t = \frac{\left|\bar{x} - \mu\right|}{\sigma / \sqrt{n-1}},\ if\ \sigma\ is\ known\ ; \qquad t = \frac{\left|\bar{x} - \mu\right|}{S / \sqrt{n-1}},\ if\ \sigma\ is\ not\ known$

$$where\ \ S = \sqrt{\frac{\sum x_i^2}{n} - \left(\bar{x}\right)^2}$$

Conclusion:

If t-calculated value < t critical value or table value at α% level of significance with *n–1* degrees of freedom then we accept H_0.

Otherwise we reject H_0

Confidence Limits or Fudicial limits: $\quad t = \bar{x} \pm t_\alpha SE\left(\bar{x}\right)$

Problems:

1. A mechanist is making engine parts with axle diameters of 0.7 inch. A random sample of 10 parts shows a mean diameter of 0.742 inch with a SD of 0.04 inch. Compute the statistic you would use to test whether the work is meeting the specification at 0.05 level of significance?
2. A sample of 26 bulbs gives a mean life of 990 hours with a SD of 20 hrs. The manufacturer claims that the mean life of bulbs is 1000 hrs. Is the sample not upto the standard.
3. A random sample of 10 boys had the following IQ's: 70, 120, 110, 101, 88, 83, 95, 98, 107 and 100.
   a) Do these data support the assumption of a population mean IQ of 100?
   b) Find a reasonable range in which most of the mean IQ values of samples of 10 boys lie.
4. Producer of gutkha claims that the nicotine content in his gutkha on the average is 1.83mg. Can this claim accepted if a random sample of 8 gutkha of this type have the nicotine contents of 2, 1.7, 2.1, 1.9, 2.2, 2.1, 2, 1.6mg? Use a0.05 level of significance.
5. A random sample of 8 envelopes is taken from the letter box of a post office and their weights in grams are found to be: 12.1, 11.9, 12.4, 12.3, 11.5, 11.6, 12.1 and 12.4
   a) Does this sample indicate at 1% level that the average weight of envelopes received at their post office is 12.35 gms.
   b) Find 95% confidence limits for the mean weight of the envelopes received at that post office.

6. Random samples of 10 bags of pesticide are taken whose weights are 50, 49, 52, 44, 45, 48, 46, 45, 49, 45 (in kgs). Test whether the average packing can be taken to be 50 kgs.

Student's t – test for Difference of two Means

Let $\bar{x}_1 \; and \; \bar{x}_2$ be the two sample means of the small sample sizes *'n₁'*and *'n₂'* drawn from two normal populations having mean $\mu_1 \; and \; \mu_2$ and standard deviation $\sigma_1 \; and \; \sigma_2$

Null Hypothesis H_0:

There is no significant difference between the two population means $\mu_1 = \mu_2$ (or)

The two samples are drawn from same population

Alternative Hypothesis H_1:

There is a significant difference between the two population means

$$\mu_1 \neq \mu_2 \; (TwoTailed)$$
$$\left.\begin{array}{l}\mu_1 < \mu_2 \\ \mu_1 > \mu_2\end{array}\right\} (OneTailed) \quad \text{(or)}$$

The two samples are not drawn from same population

Level of Significance α:

Choose the level of significance based on the problem. Default 5%

Test Statistic: SE=Standard Error

$$t = \frac{\left|\bar{x}_1 - \bar{x}_2\right|}{SE\left(\bar{x}_1 - \bar{x}_2\right)}; \sim t - distribution \; with \;\; n_1 + n_2 - 2 \deg rees \, of \;\; freedom$$

$$t = \frac{\left|\bar{x}_1 - \bar{x}_2\right|}{S\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}, \; where \; S = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}} \;\; (or);$$

$$S = \sqrt{\frac{\sum\left(x_{1i} - \bar{x}_1\right)^2 + \sum\left(x_{2i} - \bar{x}_2\right)^2}{n_1 + n_2 - 2}};$$

Conclusion:

If t-calculated value < t critical value or table value at α% level of significance with $n_1 + n_2 - 1$ degrees of freedom then we accept H_0.

Otherwise we reject H_0

Confidence Limits or Fudicial limits: $t = \left|\bar{x}_1 - \bar{x}_2\right| \pm t_\alpha SE\left(\bar{x}_1 - \bar{x}_2\right)$

Problems:

1. Below are given the gain in weights of pigs fed on two diets A and B:

Diet A	25	32	30	34	24	14	32	24	30	31	35	25			
Diet B	44	34	22	10	47	31	40	30	32	35	18	21	35	29	22

   Test if the two diets differ significantly as regards their effect on increase in weight.

2. Two horses A and B where tested according to the time (in seconds) to run a particular track with the following results:

Horse A	28	30	32	33	33	29	34
Horse B	29	30	30	24	27	29	

   Test whether the two horses has the same running capacity

3. To examine the hypothesis that the husbands are more intelligent than the wifes, an investigator took a sample of 10 couples and administered them a test which measures the IQ and the results are as follows:

Husband	117	105	97	105	123	109	86	78	103	107
Wife	106	98	87	104	116	95	90	69	108	85

   Test the hypothesis with a reasonable test at the level of significance at 0.05

4. To compare two kinds of bumper guards 6 of each kind were mounted on a car and then the car was run into a concurrence wall. The following are the costs of repairs

Guard 1	107	148	123	165	102	119
Guard 2	134	115	112	151	133	129

   Use the 0.01 the level of significance to test whether the difference between two sample means is significant.

5. Random samples of specimens of coal from two mines A and B are drawn and their heat producing capacity (in millions of calories / ton) where measured yielding the following results:

Mine A	8350	8070	8340	8130	8260	-
Mine B	7900	8140	7920	7840	7890	7950

   Is there significant difference between the means of these two samples at 0.01 level of significance.

Student's t – test for paired samples

Paired observations arise in many practical situations where each homogenous experimental unit receives both population conditions. As a result each experimental unit has a pair of observations.

Example:

To test the effectiveness of drug some 11 persons blood pressure is measured before and after the intake of certain drug.

If (x_1,y_1), (x_2,y_2), (x_3,y_3), …, (x_n,y_n) be the pairs of sales of data before and after the sales promotion in a business concern then we apply paired t- test to examine the significance of the difference of two situations.

Null Hypothesis H_0:

There is no significant difference in the performance between the before and after sample data $\mu_1 = \mu_2$ (or)

The sample is drawn from the normal population

Alternative Hypothesis H_1:

There is no significant difference in the performance between the before and after sample data

$$\mu_1 \neq \mu_2 (TwoTailed)$$
$$\left.\begin{array}{l}\mu_1 < \mu_2 \\ \mu_1 > \mu_2\end{array}\right\} (OneTailed) \text{ (or)}$$

The sample is not drawn from the normal population

Level of Significance α:

Choose the level of significance based on the problem. Default 5%

Test Statistic:

$$t = \frac{\overline{d}}{S/\sqrt{n-1}}; \quad \sim t - distribution\, with \;\; n-1 \deg rees\, of \;\; freedom$$

SE=Standard Error

$$where, \quad \overline{d} = \frac{\sum d_i}{n}; \quad d_i = x_i - y_i; \quad S = \sqrt{\frac{\sum d_i^2}{n} - \left(\overline{d}\right)^2}$$

Conclusion:

If t-calculated value < t critical value or table value at α% level of significance with *n–1* degrees of freedom then we accept H_0.

Otherwise we reject H_0

Confidence Limits or Fudicial limits: $t = \overline{d} \pm t_\alpha \frac{S}{\sqrt{n-1}}$

Problems:

1. Ten workers were given a training program with a view to study their assembly time for a certain mechanism. The result of the time and motion studies before and after the training program are given below:

Workers	1	2	3	4	5	6	7	8	9	10
X	15	18	20	17	16	14	21	19	13	22
Y	14	16	21	10	15	18	17	16	14	20

   X = Time taken for assembling before training; Y = Time taken for assembling after training;
   Test whether there is significant in assembling times before and after training.

2. Scores obtained in a shooting competition by 10 soldiers before and after intensive training are given below:

Before	67	24	57	55	63	54	56	68	33	43
After	70	38	58	58	56	67	68	75	42	38

   Test whether the intensive training is useful at 0.05 level of significance

3. The blood pressure of 5 women before and after intake of certain drug are given below:

Before	110	120	125	132	125
After	120	118	125	136	121

   Test whether there is significant change in Blood Pressure at 0.01 level of significance.

4. Memory capacity of 10 students were tested before and after training. State whether the training was effective or not from the following scores:

Before	12	14	11	8	7	10	3	0	5	6
After	15	16	10	7	5	12	10	2	3	8

5. The average losses of workers, before and after certain program are given below. Use 0.05 level of significance to test whether the program is effective (paired sample t – test). 40 and 35, 70 and 65, 45 and 42, 120 and 116, 35 and 33, 55 and 50, 77 and 73.

F-Test

When testing the significance of the difference of the means of two samples, we assumed that the two samples came from the same population or from populations with equal variances. If the variances of the populations are not equal, a significant difference in the means may arise. Hence, before we apply the t-test for the significance of the difference of two means, we have to test for the equality of population variances using F-Test of significance.

If S_1^2 and S_2^2 are the two sample variances with sample size n_1 and n_2 drawn from the normal populations.

Null Hypothesis H_0:

There is no significant difference between the two population variances $\sigma_1^2 = \sigma_2^2$ (or)

The samples are drawn from the same population

Alternative Hypothesis H_1:

There is a significant difference between the two population variances

$$\begin{array}{l} \sigma_1^2 \neq \sigma_2^2 (TwoTailed) \\ \left.\begin{array}{l} \sigma_1^2 < \sigma_2^2 \\ \sigma_1^2 > \sigma_2^2 \end{array}\right\} (OneTailed) \end{array} \text{(or)}$$

The samples are not drawn from the same population

Level of Significance α:

Choose the level of significance based on the problem. Default 5%

Test Statistic:

$$F = \frac{Greater\ Variance}{Smaller\ Variance}$$

$$F = \frac{S_1^2}{S_2^2}; \quad if\ S_1^2 > S_2^2 \sim F-distrubution\, with \left(n_1-1, n_2-1\right) \deg rees\, of\ \ freedom$$

$$F = \frac{S_2^2}{S_1^2}; \quad if\ S_2^2 > S_1^2 \sim F-distrubution\, with \left(n_2-1, n_1-1\right) \deg rees\, of\ \ freedom$$

$$where,\ S_1^2 = \frac{\sum\left(x_i - \overline{x}\right)^2}{n_1 - 1}; \quad S_2^2 = \frac{\sum\left(y_i - \overline{y}\right)^2}{n_2 - 1}$$

$\sum\left(x_i - \overline{x}\right)^2$ = the sum of squares of the deviations from the sample mean

Conclusion:

If F-calculated value < F critical value or table value at α% level of significance with (n_1-1, n_2-1) or (n_2-1, n_1-1) degrees of freedom then we accept H_0.

Otherwise we reject H_0

Problems:

1. In one sample of 8 observations from a normal population, the sum of the squares of deviations of the sample values from the sample mean is 84.4 and in another sample of 10 observations it was 102.6. Test at 5% level whether the populations have the same variance.
2. It is known that the mean diameters of rivets produced by two firms A and B are practically the same but the standard deviations may differ. For 22 rivets produced by firm A, the S.D. 2.9mm, while for 16 rivets manufactured by firm B, the SD is 3.8mm. Compute the statistic you would use to test whether the products of firm A have the same variability as those of firm B and test its significance.
3. The nicotine contents in milligrams in two samples of tobacco were found to be as follows:

Sample A	24	27	26	21	25	
Sample B	27	30	28	31	22	36

   Can it be said that the two samples have come from the same normal population.
4. Two independent samples of 8 and 7 items respectively had the following values of the variables.

Sample I	9	11	13	11	16	10	12	14
Sample II	11	13	11	14	10	8	10	--

   Do the estimates of the population variance differ significantly.
5. The measurements of the output of two units have given the following results. Assuming that both samples have been obtained from the normal populations at 10% significant level, test whether the two populations have the same variance.

Unit-A	14.1	10.1	14.7	13.7	14
Unit-B	14	14.5	13.7	12.7	14.1

CHI-SQUARE test (χ^2 – test)

When a fair coin is tossed 100 times, the theoretical considerations lead us to expect 50 heads and 50 tails. But in practice, these results are rarely achieved i.e., the results obtained in an experiment do not agree exactly with the theoretical results.

The magnitude of discrepancy between the theory and observation is given by the quantity χ^2.

If a set of events A_i *(i=1,2,…n)* are observed to occur with frequencies O_i *(i=1,2,…n)* respectively and according to probability rules A_i *(i=1,2,…n)* are expected to occur with frequencies E_i *(i=1,2,…n)* respectively with O_i *(i=1,2,…n)* are called observed frequencies and A_i *(i=1,2,…n)* are called expected frequencies.

χ^2 – test is used to test whether difference between observed and expected frequencies are significant and mainly used in:

1. Test the goodness of fit
2. Test the independence of attributes

χ^2 – TEST – THE GOODNESS OF FIT

Null Hypothesis H_0:

There is no significant difference between the observed and expected frequencies (or)

The sample is drawn from the same population

Alternative Hypothesis H_1:

There is a significant difference between the observed and expected frequencies (or)

The sample is not drawn from the same population

Level of Significance α:

Choose the level of significance based on the problem. Default 5%

Test Statistic:

$$\chi^2 = \frac{(O_i - E_i)^2}{E_i}; \sim \chi^2 - distrubution\, with\, (n-1) \deg rees\, of\ \ freedom$$

Expected frequencies will be computed as the average value of the observed frequencies or the probability value or a criteria specified in the problem.

Conclusion:

If χ^2-calculated value < χ^2 critical value or table value at α% level of significance with (n –1) degrees of freedom then we accept H_0.

Otherwise we reject H_0

Problems:

1. The number of automobile accidents per week in a certain community are as follows: 12, 8, 20, 2, 14, 10, 15, 6, 9, 4. Are these frequencies in agreement with the belief that accident conditions were the same during this 10 week period.
2. A sample analysis of examination results of 500 students were made, it was found that 220 students has failed, 170 ha secured third class, 90 were placed in second class and 20 got first class. Do these figure commensurate with the general examination result which is in the ratio of 4:3:2:1 for the various categories respectively.
3. A pair of dice are thrown 360 times and the frequency of each sum is indicated below:

Sum	2	3	4	5	6	7	8	9	10	11	12
Frequency	8	24	35	37	44	65	51	42	26	14	14

   Would you say that the dice are fair on the basis of the Chi-square test at 0.05 level of significance
4. 4 coins were tossed 160 times and the following results were obtained.

No of Heads	0	1	2	3	4
Observed frequencies	17	52	54	31	6

   Under the assumption that coins are balanced, find the expected frequencies of 0, 1, 2, 3, or 4 heads and test the goodness of fit.
5. Fit a Poisson distribution to the following data and for its goodness of fit at 0.05

x	0	1	2	3	4
f	419	352	154	56	19

χ^2 – TEST – THE INDEPENDENCE OF ATTRIBUTES

Attribute: Literally, an attribute means a quality or characteristic.

Ex: Drinking, Smoking, Blindness, Honesty, Beauty, etc.

An attribute may be marked by its presence or absence in a number of a given population. Let the observations be classified according to two attributes and the frequencies O_i in the different categories be shown in a two-way table called contingency table.

To test the two attributes are independent we apply chi-square test:

Null Hypothesis H_0:

There is no significant difference between the observed and expected frequencies (or)

The two attributes are independent

Alternative Hypothesis H_1:

There is a significant difference between the observed and expected frequencies (or)

The two attributes are dependent

Level of Significance α:

Choose the level of significance based on the problem. Default 5%

Test Statistic:

$$\text{Corresponding Expected frequency} = \frac{Row Total * Column\ Total}{Grand\ Total}$$

Sample: Let A, B, C, D be the four attributes then a, b, c, d are observed frequencies and a_1, b_1, c_1, d_1 are expected frequencies then the contingency table is:

Observed frequencies			
	A	**B**	**Total**
C	a	B	**a+b**
D	c	D	**c+d**
Total	**a+c**	**b+d**	**N=a+b+c+d**

Expected frequencies			
	A	**B**	**Total**
C	$a_1 = \frac{(a+b)(a+c)}{N}$	$b_1 = \frac{(a+b)(b+d)}{N}$	$\mathbf{a_1+b_1}$
D	$c_1 = \frac{(c+d)(a+c)}{N}$	$d_1 = \frac{(c+d)(b+d)}{N}$	$\mathbf{c_1+d_1}$
Total	$\mathbf{a_1+c_1}$	$\mathbf{b_1+d_1}$	**N**

$$\chi^2 = \frac{(O_i - E_i)^2}{E_i};\ \sim \chi^2 - distrubution\, with \left[(m-1).(n-1)\right] \deg rees\, of\ \ freedom$$

m = number of rows; n = number of columns in the contingency table

Conclusion:

If χ^2-calculated value $<$ χ^2 critical value or table value at α% level of significance with (m–1, n –1) degrees of freedom then we accept H_0.

Otherwise we reject H_0

Problems:

1. On the basis of the information given below about the treatment of 200 patients suffering from a disease state whether the new treatment is comparatively superior to the conventional treatment.

	Favourable	Not Favourable
New	60	30
Conventional	40	70

2. Given the following contingency table for hair colour and eye colour. Find the value of Chi-square. Is there good association between the two?

Eye Colour	Hair Colour		
	Fair	Brown	Black
Blue	15	5	20
Grey	20	10	20
Brown	25	15	20

3. Two researchers adopted different sampling techniques while investigating some group of students to find the number of students falling into different intelligence level. The results are as follows:

Researchers	Below average	Average	Above Average	Genius
X	86	60	44	10
Y	40	33	25	2

Would you say that the sampling techniques adopted by the two researchers are significantly different?

4. From the following data, find whether there is any significant liking in the habit of taking soft drinks among the categories of employees.

Soft drinks	Clerks	Teachers	Officers
Pepsi	10	25	65
Thumsup	15	30	65
Fanta	50	60	30

SAMPLING TECHNIQUES

Basic concepts of sampling

Essentially, sampling consists of obtaining information from only a part of a large group or population so as to infer about the whole population. The object of sampling is thus to secure a sample which will represent the population and reproduce the important characteristics of the population under study as closely as possible.

The principal advantages of sampling as compared to complete enumeration of the population are reduced cost, greater speed, greater scope and improved accuracy. Many who insist that the only accurate way to survey a population is to make a complete enumeration, overlook the fact that there are many sources of errors in a complete enumeration and that a hundred per cent enumeration can be highly erroneous as well as nearly impossible to achieve. In fact, a sample can yield more accurate results because the sources of errors connected with reliability and training of field workers, clarity of instruction, mistakes in measurement and recording, badly kept measuring instruments, misidentification of sampling units, biases of the enumerators and mistakes in the processing and analysis of the data can be controlled more effectively. The smaller size of the sample makes the supervision more effective. Moreover, it is important to note that the precision of the estimates obtained from certain types of samples can be estimated from the sample itself. The net effect of a sample survey as compared to a complete enumeration is often a more accurate answer achieved with fewer personnel and less work at a low cost in a short time.

The most ‘convenient’ method of sampling is that in which the investigator selects a number of sampling units which he considers ‘representative’ of the whole population. For example, in estimating the whole volume of a forest stand, he may select a few trees which may appear to be of average dimensions and typical of the area and measure their volume. A walk over the forest area with an occasional stop and flinging a stone with the eyes closed or some other simple way that apparently avoids any deliberate choice of the sampling units is very tempting in its simplicity. However, it is clear that such methods of selection are likely to be biased by the investigator’s judgement and the results will thus be biased and unreliable. Even if the investigator can be trusted to be completely objective, considerable conscious or unconscious errors of judgement, not frequently recognized, may occur and such errors due to bias may far outweigh any supposed increase in accuracy resulting from deliberate or purposive selection of the units. Apart from the above points, subjective sampling does not permit the evaluation of the precision of the estimates calculated from samples. Subjective sampling is statistically unsound and should be discouraged.

When sampling is performed so that every unit in the population has some chance of being selected in the sample and the probability of selection of every unit is known, the method of sampling is called probability sampling. An example of probability sampling is random selection, which should be clearly distinguished from haphazard selection, which implies a strict process of selection equivalent to that of drawing lots. In this manual, any reference to sampling, unless otherwise stated, will relate to some form of probability sampling. The probability that any sampling unit will be selected in the sample depends on the sampling procedure used. The important point to note is that the precision and reliability of the estimates obtained from a sample can be evaluated only for a probability sample. Thus the errors of sampling can be controlled satisfactorily in this case.

The object of designing a sample survey is to minimise the error in the final estimates. Any forest survey involving data collection and analysis of the data is subject to a variety of errors. The errors may be classified into two groups *viz*., (i) non-sampling errors (ii) sampling errors. The non-sampling errors like the errors in location of the units, measurement of the characteristics, recording mistakes, biases of enumerators and faulty methods of analysis may contribute substantially to the total error of the final results to both complete enumeration and sample surveys. The magnitude is likely to be larger in complete enumeration since the smaller size of the sample project makes it possible to be more selective in assignment of personnel for the survey operations, to be more thorough in their training and to be able to concentrate to a much greater degree on reduction of non-sampling errors. Sampling errors arise from the fact that only a fraction of the forest area is enumerated. Even if the sample is a probability sample, the sample being based on observations on a part of the population cannot, in general, exactly represent the population. The average magnitude of the sampling errors of most of the probability samples can be estimated from the data collected. The magnitude of the sampling errors, depends on the size of the sample, the variability within the population and the sampling method adopted. Thus if a probability sample is used, it is possible to predetermine the size of the sample needed to obtain desired and specified degree of precision.

A sampling scheme is determined by the size of sampling units, number of sampling units to be used, the distribution of the sampling units over the entire area to be sampled, the type and method of measurement in the selected units and the statistical procedures for analysing the survey data. A variety of sampling methods and estimating techniques developed to meet the varying demands of the survey statistician accord the user a wide selection for specific situations. One can choose the method or combination of methods that will yield a desired degree of precision at minimum cost. Additional references are Chacko (1965) and Sukhatme *et al, (1984)*

The principal steps in a sample survey

In any sample survey, we must first decide on the type of data to be collected and determine how adequate the results should be. Secondly, we must formulate the sampling plan for each of the characters for which data are to be collected. We must also know how to combine the sampling procedures for the various characters so that no duplication of field work occurs. Thirdly, the field work must be efficiently organised with adequate provision for supervising the work of the field staff. Lastly, the analysis of the data collected should be carried out using appropriate statistical techniques and the report should be drafted giving full details of the basic assumptions made, the sampling plan and the results of the statistical analysis. The report should contain estimate of the margin of the sampling errors of the results and may also include the possible effects of the non-sampling errors. Some of these steps are elaborated further in the following.

(i) Specification of the objectives of the survey: Careful consideration must be given at the outset to the purposes for which the survey is to be undertaken. For example, in a forest survey, the area

to be covered should be decided. The characteristics on which information is to be collected and the degree of detail to be attempted should be fixed. If it is a survey of trees, it must be decided as to what species of trees are to be enumerated, whether only estimation of the number of trees under specified diameter classes or, in addition, whether the volume of trees is also proposed to be estimated. It must also be decided at the outset what accuracy is desired for the estimates.

(*ii*) *Construction of a frame of units :* The first requirement of probability sample of any nature is the establishment of a frame. The structure of a sample survey is determined to a large extent by the frame. A frame is a list of sampling units which may be unambiguously defined and identified in the population. The sampling units may be compartments, topographical sections, strips of a fixed width or plots of a definite shape and size.

The construction of a frame suitable for the purposes of a survey requires experience and may very well constitute a major part of the work of planning the survey. This is particularly true in forest surveys since an artificial frame composed of sampling units of topographical sections, strips or plots may have to be constructed. For instance, the basic component of a sampling frame in a forest survey may be a proper map of the forest area. The choice of sampling units must be one that permits the identification in the field of a particular sampling unit which has to be selected in the sample. In forest surveys, there is considerable choice in the type and size of sampling units. The proper choice of the sampling units depends on a number of factors; the purpose of the survey, the characteristics to be observed in the selected units, the variability among sampling units of a given size, the sampling design, the field work plan and the total cost of the survey. The choice is also determined by practical convenience. For example, in hilly areas it may not be practicable to take strips as sampling units. Compartments or topographical sections may be more convenient. In general, at a given intensity of sampling (proportion of area enumerated) the smaller the sampling units employed the more representative will be the sample and the results are likely to be more accurate.

(iii) Choice of a sampling design: If it is agreed that the sampling design should be such that it should provide a statistically meaningful measure of the precision of the final estimates, then the sample should be a probability sample, in that every unit in the population should have a known probability of being selected in the sample. The choice of units to be enumerated from the frame of units should be based on some objective rule which leaves nothing to the opinion of the field worker. The determination of the number of units to be included in the sample and the method of selection is also governed by the allowable cost of the survey and the accuracy in the final estimates.

(iv) Organisation of the field work : The entire success of a sampling survey depends on the reliability of the field work. In forest surveys, the organization of the field work should receive the utmost attention, because even with the best sampling design, without proper organization the sample results may be incomplete and misleading. Proper selection of the personnel, intensive training, clear instructions and proper supervision of the fieldwork are essential to obtain satisfactory results. The field parties should correctly locate the selected units and record the necessary measurements according to the specific instruction given. The supervising staff should check a part of their work in the field and satisfy that the survey carried out in its entirety as planned.

(v) Analysis of the data : Depending on the sampling design used and the information collected, proper formulae should be used in obtaining the estimates and the precision of the estimates should be computed. Double check of the computations is desired to safeguard accuracy in the analysis.

(vi) Preliminary survey (pilot trials) : The design of a sampling scheme for a forest survey requires both knowledge of the statistical theory and experience with data regarding the nature of the forest area, the pattern of variability and operational cost. If prior knowledge in these matters is not available, a statistically planned small scale 'pilot survey' may have to be conducted before undertaking any large scale survey in the forest area. Such exploratory or pilot surveys will provide adequate knowledge regarding the variability of the material and will afford opportunities to test and improve field procedures, train field workers and study the operational efficiency of a design. A pilot survey will also provide data for estimating the various components of cost of operations in a survey like time of travel, time of location and enumeration of sampling units, etc. The above information will be of great help in deciding the proper type of design and intensity of sampling that will be appropriate for achieving the objects of the survey.

Sampling terminology

Although the basic concepts and steps involved in sampling are explained above, some of the general terms involved are further clarified in this section so as to facilitate the discussion on individual sampling schemes dealt with in later sections.

Population : The word population is defined as the aggregate of units from which a sample is chosen. If a forest area is divided into a number of compartments and the compartments are the units of sampling, these compartments will form the population of sampling units. On the other hand, if the forest area is divided into, say, a thousand strips each 20 m wide, then the thousand strips will form the population. Likewise if the forest area is divided into plots of, say, one-half hectare each, the totality of such plots is called the population of plots.

Sampling units : Sampling units may be administrative units or natural units like topographical sections and subcompartments or it may be artificial units like strips of a certain width, or plots of a definite shape and size. The unit must be a well defined element or group of elements identifiable in the forest area on which observations on the characteristics under study could be made. The population is thus sub-divided into suitable units for the purpose of sampling and these are called sampling units.

Sampling frame : A list of sampling units will be called a 'frame'. A population of units is said to be finite if the number units in it is finite.

Sample : One or more sampling units selected from a population according to some specified procedure will constitute a sample.

Sampling intensity : Intensity of sampling is defined as the ratio of the number of units in the sample to the number of units in the population.

Population total : Suppose a finite population consists of units $U_1, U_2, \ldots, U_N$. Let the value of the characteristic for the *i*th unit be denoted by y_i. For example the units may be strips and the characteristic may be the number of trees of a certain species in a strip. The total of the values y_i ($i = 1, 2, \ldots, N$), namely,

$$Y = \sum_{i=1}^{N} y_i \quad (5.1)$$

is called the population total which in the above example is the total number of trees of the particular species in the population.

Population mean : The arithmetic mean

$$\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} y_i \quad (5.2)$$

is called the population mean which, in the example considered, is the average number of trees of the species per strip.

Population variance : A measure of the variation between units of the population is provided by the population variance

$$S_y^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{Y})^2 = \frac{1}{N}\sum_{i=1}^{N} y_i^2 - \bar{Y}^2 \quad (5.3)$$

which in the example considered measures the variation in number of trees of the particular species among the strips. Large values of the population variance indicate large variation between units in the population and small values indicate that the values of the characteristic for the units are close to the population mean. The square root of the variance is known as *standard deviation.*

Coefficient of variation : The ratio of the standard deviation to the value of the mean is called the coefficient of variation, which is usually expressed in percentage.

$$C.\ V. = \frac{S_y}{\bar{Y}} \quad (5.4)$$

The coefficient of variation, being dimensionless, is a valuable tool to compare the variation between two or more populations or sets of observations.

Parameter : A function of the values of the units in the population will be called a parameter. The population mean, variance, coefficient of variation, etc., are examples of population parameters. The problem in sampling theory is to estimate the parameters from a sample by a procedure that makes it possible to measure the precision of the estimates.

Estimator, estimate : Let us denote the sample observations of size n by $y_1, y_2, \ldots, y_n$. Any function of the sample observations will be called a *statistic*. When a statistic is used to estimate a population parameter, the statistic will be called an estimator. For example, the sample mean is an estimator of the population mean. Any particular value of an estimator computed from an observed sample will be called an estimate.

Bias in estimation : A statistic t is said to be an unbiased estimator of a population parameter q if its expected value, denoted by E(t), is equal to q . A sampling procedure based on a probability scheme gives rise to a number of possible samples by repetition of the sampling procedure. If the values of the statistic t are computed for each of the possible samples and if the average of the values is equal to the population value q , then t is said to be an unbiased estimator of q based on sampling procedure. Notice that the repetition of the procedure and computing the values of t for each sample is only conceptual, not actual, but the idea of generating all possible estimates by repetition of the sampling process is fundamental to the study of bias and of the assessment of sampling error. In case E(t) is not equal to q , the statistic t is said to be a biased estimator of q and the bia s is given by, bias = E(t) - q . The introduction of a genuinely random process in selecting a sample is an important step in avoiding bias. Samples selected subjectively will usually be very seriously biased. In forest surveys, the tendency of forest officers to select typical forest areas for enumerations, however honest the intention may be, is bound to result in biased estimates.

Sampling variance : The difference between a sample estimate and the population value is called the sampling error of the estimate, but this is naturally unknown since the population value is unknown. Since the sampling scheme gives rise to different possible samples, the estimates will differ from sample to sample. Based on these possible estimates, a measure of the average magnitude over all possible samples of the squares of the sampling error can be obtained and is known as the *mean square error* (*MSE*) of the estimate which is essentially a measure of the divergence of an estimator from the true population value. Symbolically, $MSE = E[t - q]^2$. The sampling variance ($V(t)$)is a measure of the divergence of the estimate from its expected value. It is defined as the average magnitude over all possible samples of the squares of deviations of the estimator from its expected value and is given by $V(t) = E[t - E(t)]^2$.

Notice that the sampling variance coincides with the mean square error when t is an unbiased estimator. Generally, the magnitude of the estimate of the sampling variance computed from a sample is taken as indicating whether a sample estimate is useful for the purpose. The larger the sample and the smaller the variability between units in the population, the smaller will be the sampling error and the greater will be the confidence in the results.

Standard error of an estimator : The square root of the sampling variance of an estimator is known as the standard error of the estimator. The standard error of an estimate divided by the value of the estimate is called relative standard error which is usually expressed in percentage.

Accuracy and precision : The standard error of an estimate, as obtained from a sample, does not include the contribution of the bias. Thus we may speak of the standard error or the sampling variance of the estimate as measuring on the inverse scale, the precision of the estimate, rather than its *accuracy.* Accuracy usually refers to the size of the deviations of the sample estimate from the mean *m* = E (*t*) obtained by repeated application of the sampling procedure, the bias being thus measured by *m* - q .

It is the accuracy of the sample estimate in which we are chiefly interested; it is the precision with which we are able to measure in most instances. We strive to design the survey and attempt to analyse the data using appropriate statistical methods in such a way that the precision is increased to the maximum and bias is reduced to the minimum.

Confidence limits : If the estimator *t* is normally distributed (which assumption is generally valid for large samples), a confidence interval defined by a lower and upper limit can expected to include the population parameter q with a specified probability level. The limits are given by

Lower limit = $t - z\sqrt{\hat{V}(t)}$ (5.5)

Upper limit = $t + z\sqrt{\hat{V}(t)}$ (5.6)

where $\hat{V}(t)$ is the estimate of the variance of *t* and *z* is the value of the normal deviate corresponding to a desired P % confidence probability. For example, when *z* is taken as 1.96, we say that the chance of the true value of q being contained in the random interval defined by the lower and upper confidence limits is 95 per cent. The confidence limits specify the range of variation expected in the population mean and also stipulate the degree of confidence we should place in our sample results. If the sample size is less than 30, the value of *k* in the formula for the lower and upper confidence limits should be taken from the percentage points of Student's *t* distribution (See Appendix 2) with degrees of freedom of the sum of squares in the estimate of the variance of *t*. Moderate departures of the distribution from normality does not affect appreciably the formula for the confidence limits. On the other hand, when the distribution is very much different from normal, special methods are needed. For example, if we use small area sampling units to estimate the average number of trees in higher diameter classes, the distribution may have a large skewness. In such cases, the above formula for calculating the lower and upper confidence limits may not be directly applicable.

Some general remarks : In the sections to follow, capital letters will usually be used to denote population values and small letters to denote sample values. The symbol 'cap' (^) above a symbol for a population value denotes its estimate based on sample observations. Other special notations used will be explained as and when they are introduced.

While describing the different sampling methods below, the formulae for estimating only population mean and its sampling variance are given. Two related parameters are population total and ratio of the character under study (*y*) to some auxiliary variable (*x*). These related statistics can always be obtained from the mean by using the following general relations.

$$\hat{Y} = N\hat{\bar{Y}} \quad (5.7)$$

$$V(\hat{Y}) = N^2 V(\hat{\bar{Y}}) \quad (5.8)$$

$$\hat{R} = \frac{\hat{Y}}{X} \quad (5.9)$$

$$V(\hat{R}) = \frac{V(\hat{Y})}{X^2} \quad (5.10)$$

where $\hat{Y}$ = Estimate of the population total

N = Total number of units in the population

$\hat{R}$ = Estimate of the population ratio

X = Population total of the auxiliary variable

Simple random sampling

A sampling procedure such that each possible combination of sampling units out of the population has the same chance of being selected is referred to as simple random sampling. From theoretical considerations, simple random sampling is the simplest form of sampling and is the basis for many other sampling methods. Simple random sampling is most applicable for the initial survey in an investigation and for studies which involve sampling from a small area where the sample size is relatively small. When the investigator has some knowledge regarding the population sampled, other methods which are likely to be more efficient and convenient for organising the survey in the field, may be adopted. The irregular distribution of the sampling units in the forest area in simple random sampling may be of great disadvantage in forest areas where accessibility is poor and the costs of travel and locating the plots are considerably higher than the cost of enumerating the plot.

Selection of sampling units

In practice, a random sample is selected unit by unit. Two methods of random selection for simple random sampling without replacement are explained in this section.

(i) Lottery method : The units in the population are numbered 1 to *N*. If *N* identical counters with numberings 1 to *N* are obtained and one counter is chosen at random after shuffling the counters, then the probability of selecting any counter is the same for all the counters. The process is repeated *n* times without replacing the counters selected. The units which correspond to the numbers on the chosen counters form a simple random sample of size *n* from the population of *N* units.

(ii) Selection based on random number tables : The procedure of selection using the lottery method, obviously becomes rather inconvenient when N is large. To overcome this difficulty, we may use a table of random numbers such as those published by Fisher and Yates (1963) a sample of which is given in Appendix 6. The tables of random numbers have been developed in such a way that the digits 0 to 9 appear independent of each other and approximately equal number of times in the table. The simplest way of selecting a random sample of required size consists in selecting a set of n random numbers one by one, from 1 to N in the random number table and, then, taking the units bearing those numbers. This procedure may involve a number of rejections since all the numbers more than N appearing in the table are not considered for selection. In such cases, the procedure is modified as follows. If N is a d digited number, we first determine the highest d digited multiple of N, say N'. Then a random number r is chosen from 1 to N' and the unit having the serial number equal to the remainder obtained on dividing r by N, is considered as selected. If remainder is zero, the last unit is selected. A numerical example is given below.

Suppose that we are to select a simple random sample of 5 units from a serially numbered list of 40 units. Consulting Appendix 6 : Table of random numbers, and taking column (5) containing two-digited numbers, the following numbers are obtained:

39, 27, 00, 74, 07

In order to give equal chances of selection to all the 100 units, we are to reject all numbers above 79 and consider (00) equivalent to 80. We now divide the above numbers in turn by 40 and take the remainders as the selected strip numbers for our sample, rejecting the remainders that are repeated. We thus get the following 16 strip numbers as our sample :

39, 27, 40, 34, 7.

Parameter estimation

Let $y_1, y_2, \ldots, y_n$ be the measurements on a particular characteristic on n selected units in a sample from a population of N sampling units. It can be shown in the case of simple random sampling without replacement that the sample mean,

$$\hat{\bar{Y}} = \bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} \quad (5.11)$$

is an unbiased estimator of the population mean, $\bar{Y}$. An unbiased estimate of the sampling variance of $\bar{y}$ is given by

$$\hat{V}(\hat{\bar{Y}}) = \frac{N-n}{Nn} s_y^2 \quad (5.12)$$

where $s_y^2 = \dfrac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}$ (5.13)

Assuming that the estimate $\bar{y}$ is normally distributed, a confidence interval on the population mean $\bar{Y}$ can be set with the lower and upper confidence limits defined by,

Lower limit $\hat{\bar{Y}}_L = \bar{y} - z\dfrac{s_y}{\sqrt{n}}\sqrt{\dfrac{N-n}{N}}$ (5.14)

Upper limit $\hat{\bar{Y}}_U = \bar{y} + z\dfrac{s_y}{\sqrt{n}}\sqrt{\dfrac{N-n}{N}}$ (5.15)

where *z* is the table value which depends on how many observations there are in the sample. If there are 30 or more observations we can read the values from the table of the normal distribution (Appendix 1). If there are less than 30 observations, the table value should be read from the table of *t* distribution (Appendix 2), using *n* - 1 degree of freedom.

The computations are illustrated with the following example. Suppose that a forest has been divided up into 1000 plots of 0.1 hectare each and a simple random sample of 25 plots has been selected. For each of these sample plots the wood volumes in m^3 were recorded. The wood volumes were,

7 10 7 4 7

8 8 8 7 5

2 6 9 7 8

6 7 11 8 8

7 3 8 7 7

If the wood volume on the *i*th sampling unit is designated as y_i , an unbiased estimator of the population mean, is obtained using Equation (5.11) as,

$$\hat{\bar{Y}} = \bar{y} = \frac{7+8+2+\ ...\ +7}{25} = \frac{175}{25}$$

$= 7\ m^3$

which is the mean wood volume per plot of 0.1 ha in the forest area.

An estimate (s_y^2) of the variance of individual values of *y* is obtained using Equation (5.13).

$$s_y^2 = \frac{(7-7)^2 + (8-7)^2 + \ldots + (7-7)^2}{25-1}$$

$$= \frac{82}{24} = 3.833$$

Then unbiased estimate of sampling variance of is

$$\hat{V}(\hat{\bar{Y}}) = \left(\frac{1000-25}{(1000)(25)}\right) 3.833$$

$= 0.1495\ (m^3)^2$

$$SE(\hat{\bar{Y}}) = \sqrt{0.1495} = 0.3867\ m^3$$

The relative standard error which is $\frac{SE(\hat{\bar{Y}})}{\hat{\bar{Y}}}(100)$ is a more common expression. Thus,

$$RSE(\hat{\bar{Y}}) = \frac{\sqrt{0.1495}}{7}(100) = 5.52\ \%$$

The confidence limits on the population mean are obtained using Equations (5.14) and (5.15).

Lower limit $\hat{\bar{Y}}_L = 7 - (2.064)\sqrt{0.1495}$

= 6.20 cords

Upper limit $\hat{\bar{Y}}_U = 7 + (2.064)\sqrt{0.1495}$

= 7.80 cords

The 95% confidence interval for the population mean is (6.20, 7.80) m^3. Thus, we are 95% confident that the confidence interval (6.20, 7.80) m^3 would include the population mean.

An estimate of the total wood volume in the forest area sampled can easily be obtained by multiplying the estimate of the mean by the total number of plots in the population. Thus,

$\hat{Y} = 7(1000) = 7000 \text{ m}^3$

with a confidence interval of (6200, 7800) obtained by multiplying the confidence limits on the mean by $N = 1000$. The RSE of $\hat{Y}$, however, will not be changed by this operation.

Systematic sampling

Systematic sampling employs a simple rule of selecting every *k*th unit starting with a number chosen at random from 1 to *k* as the random start. Let us assume that *N* sampling units in the population are numbered 1 to *N*. To select a systematic sample of *n* units, we take a unit at random from the first *k* units and then every *k*th sampling unit is selected to form the sample. The constant *k* is known as the *sampling interval* and is taken as the integer nearest to N / n, the inverse of the sampling fraction. Measurement of every *k*th tree along a certain compass bearing is an example of systematic sampling. A common sampling unit in forest surveys is a narrow strip at right angles to a base line and running completely across the forest. If the sampling units are strips, then the scheme is known as systematic sampling by strips. Another possibility is known as systematic line plot sampling where plots of a fixed size and shape are taken at equal intervals along equally spaced parallel lines. In the latter case, the sample could as well be systematic in two directions.

Systematic sampling certainly has an intuitive appeal, apart from being easier to select and carry out in the field, through spreading the sample evenly over the forest area and ensuring a certain amount of representation of different parts of the area. This type of sampling is often convenient in exercising control over field work. Apart from these operational considerations, the procedure of systematic sampling is observed to provide estimators more efficient than simple random sampling under normal forest conditions. The property of the systematic sample in spreading the sampling units evenly over the population can be taken advantage of by listing the units so that homogeneous units are put together or such that the values of the characteristic for the units are in ascending or descending order of magnitude. For example, knowing the fertility trend of the forest area the units (for example strips) may be listed along the fertility trend.

If the population exhibits a regular pattern of variation and if the sampling interval of the systematic sample coincides with this regularity, a systematic sample will not give precise estimates. It must, however, be mentioned that no clear case of periodicity has been reported in a forest area. But the fact that systematic sampling may give poor precision when unsuspected periodicity is present should not be lost sight of when planning a survey.

Selection of a systematic sample

To illustrate the selection of a systematic sample, consider a population of $N = 48$ units. A sample of $n = 4$ units is needed. Here, $k = 12$. If the random number selected from the set of numbers from 1 to 12 is 11, then the units associated with serial numbers 11, 23, 35 and 47 will be selected. In situations where N is not fully divisible by n, k is calculated as the integer nearest to N / n. In this situation, the sample size is not necessarily n and in some cases it may be n -1.

5.3.2. Parameter estimatio n

The estimate for the population mean per unit is given by the sample mean

$$\hat{\bar{Y}} = \bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} \quad (5.16)$$

where n is the number of units in the sample.

In the case of systematic strip surveys or, in general, any one dimensional systematic sampling, an approximation to the standard error may be obtained from the differences between pairs of successive units. If there are n units enumerated in the systematic sample, there will be (n-1) differences. The variance per unit is therefore, given by the sum of squares of the differences divided by twice the number of differences. Thus if $y_1, y_2, \ldots, y_n$ are the observed values (say volume) for the n units in the systematic sample and defining the first difference $d(y_i)$ as given below,

$$d(y_i) = y_{(i+1)} - y_{(i)}; \; (i = 1, 2, \ldots, n -1), \quad (5.17)$$

the approximate variance per unit is estimated as

$$\hat{V}(\hat{\bar{Y}}) = \frac{1}{2n(n-1)} \sum_{i=1}^{n-1} \left[d(y_i) \right]^2 \quad (5.18)$$

As an example, Table 5.1 gives the observed diameters of 10 trees selected by systematic selection of 1 in 20 trees from a stand containing 195 trees in rows of 15 trees. The first tree was selected as the 8th tree from one of the outside edges of the stand starting from one corner and the remaining trees were selected systematically by taking every 20th tree switching to the nearest tree of the next row after the last tree in any row is encountered.

Table 5.1. Tree diameter recorded on a systematic sample of

10 trees from a plot.

Selected tree number	Diameter at breast-height(cm) y_i	First difference $d(y_i)$
8	14.8	
28	12.0	-2.8
48	13.6	+1.6
68	14.2	+0.6
88	11.8	-2.4
108	14.1	+2.3
128	11.6	-2.5
148	9.0	-2.6
168	10.1	+1.1
188	9.5	-0.6

Average diameter is equal to

$$\hat{\bar{Y}} = \frac{1}{10}(14.8 + 12.0 + \ldots\ldots + 9.5) = 12.07$$

The nine first differences can be obtained as shown in column (3) of the Table 5.1. The error variance of the mean per unit is thus

$$\hat{V}(\hat{\bar{Y}}) = \frac{(-2.8)^2 + (1.6)^2 + \ldots + (-0.6)^2}{2x9x10} = \frac{36.9}{180}$$

= 0.202167

A difficulty with systematic sampling is that one systematic sample by itself will not furnish valid assessment of the precision of the estimates. With a view to have valid estimates of the precision, one may resort to partially systematic samples. A theoretically valid method of using the idea of systematic samples and at the same time leading to unbiased estimates of the sampling error is to draw a minimum of two systematic samples with independent random starts.

If $\bar{y}_1$, $\bar{y}_2$, …, $\bar{y}_m$ are *m* estimates of the population mean based on *m* independent systematic samples, the combined estimate is

$$\bar{y} = \frac{1}{m}\sum_{i=1}^{m}\bar{y}_i \qquad (5.19)$$

The estimate of the variance of $\bar{y}$ is given by

$$\hat{V}(\bar{y}) = \frac{1}{m(m-1)}\sum_{i=1}^{m}(\bar{y}_i - \bar{y})^2 \qquad (5.20)$$

Notice that the precision increases with the number of independent systematic samples.

As an example, consider the data given in Table 5.1 along with another systematic sample selected with independent random starts. In the second sample, the first tree was selected as the 10th tree. Data for the two independent samples are given in Table 5.2.

Table 5.2. Tree diameter recorded on two independent systematic samples of 10 trees from a plot.

Sample 1		Sample 2	
Selected tree number	Diameter at breast-height(cm) y_i	Selected tree number	Diameter at breast-height(cm) y_i
8	14.8	10	13.6
28	12.0	30	10.0
48	13.6	50	14.8
68	14.2	70	14.2
88	11.8	90	13.8
108	14.1	110	14.5
128	11.6	130	12.0
148	9.0	150	10.0
168	10.1	170	10.5

188	9.5	190	8.5

The average diameter for the first sample is $\bar{y}_1 = 12.07$. The average diameter for the first sample is $\bar{y}_2 = 12.19$. Combined estimate of population mean ($\bar{y}$) is obtained by using Equation (5.19) as,

$$\bar{y} = \frac{1}{2}(12.07 + 12.19)$$

= 12.13

The estimate of the variance of is obtained by using Equation (5.20).

$$\hat{V}(\bar{y}) = \frac{1}{2(2-1)}(12.07 - 12.13)^2(12.19 - 12.13)^2 = 0.0036$$

$$SE(\bar{y}) = \sqrt{0.0036} = 0.06$$

One additional variant of systematic sampling is that sampling may as well be systematic in two directions. For example, in plantations, a systematic sample of rows and measurements on every tenth tree in each selected row may be adopted with a view to estimate the volume of the stand. In a forest survey, one may take a series of equidistant parallel strips extending over the whole width of the forest and the enumeration in each strip may be done by taking a systematic sample of plots or trees in each strip. Forming rectangular grids of (p x q) metres and selecting a systematic sample of rows and columns with a fixed size plot of prescribed shape at each intersection is another example.

In the case of two dimensional systematic sample, a method of obtaining the estimates and approximation to the sampling error is based on stratification and the method is similar to the stratified sampling given in section 5.4. For example, the sample may be arbitrarily divided into sets of four in 2 x 2 units and each set may be taken to form a stratum with the further assumption that the observations within each stratum are independently and randomly chosen. With a view to make border adjustments, overlapping strata may be taken at the boundaries of the forest area.

Stratified sampling

The basic idea in stratified random sampling is to divide a heterogeneous population into sub-populations, usually known as strata, each of which is internally homogeneous in which case a precise estimate of any stratum mean can be obtained based on a small sample from that stratum and by combining such estimates, a precise estimate for the whole population can be obtained.

Stratified sampling provides a better cross section of the population than the procedure of simple random sampling. It may also simplify the organisation of the field work. Geographical proximity is sometimes taken as the basis of stratification. The assumption here is that geographically contiguous areas are often more alike than areas that are far apart. Administrative convenience may also dictate the basis on which the stratification is made. For example, the staff already available in each range of a forest division may have to supervise the survey in the area under their jurisdiction. Thus, compact geographical regions may form the strata. A fairly effective method of stratification is to conduct a quick reconnaissance survey of the area or pool the information already at hand and stratify the forest area according to forest types, stand density, site quality etc. If the characteristic under study is known to be correlated with a supplementary variable for which actual data or at least good estimates are available for the units in the population, the stratification may be done using the information on the supplementary variable. For instance, the volume estimates obtained at a previous inventory of the forest area may be used for stratification of the population.

In stratified sampling, the variance of the estimator consists of only the 'within strata' variation. Thus the larger the number of strata into which a population is divided, the higher, in general, the precision, since it is likely that, in this case, the units within a stratum will be more homogeneous. For estimating the variance within strata, there should be a minimum of 2 units in each stratum. The larger the number of strata the higher will, in general, be the cost of enumeration. So, depending on administrative convenience, cost of the survey and variability of the characteristic under study in the area, a decision on the number of strata will have to be arrived at.

Allocation and selection of the sample within strata

Assume that the population is divided into k strata of $N_1, N_2, \ldots, N_k$ units respectively, and that a sample of n units is to be drawn from the population. The problem of allocation concerns the choice of the sample sizes in the respective strata, *i.e.*, how many units should be taken from each stratum such that the total sample is n.

Other things being equal, a larger sample may be taken from a stratum with a larger variance so that the variance of the estimates of strata means gets reduced. The application of the above principle requires advance estimates of the variation within each stratum. These may be available from a previous survey or may be based on pilot surveys of a restricted nature. Thus if this information is available, the sampling fraction in each stratum may be taken proportional to the standard deviation of each stratum.

In case the cost per unit of conducting the survey in each stratum is known and is varying from stratum to stratum an efficient method of allocation for minimum cost will be to take large samples from the stratum where sampling is cheaper and variability is higher. To apply this procedure one needs information on variability and cost of observation per unit in the different strata.

Where information regarding the relative variances within strata and cost of operations are not available, the allocation in the different strata may be made in proportion to the number of units in them or the total area of each stratum. This method is usually known as ‘proportional allocation’.

For the selection of units within strata, In general, any method which is based on a probability selection of units can be adopted. But the selection should be independent in each stratum. If independent random samples are taken from each stratum, the sampling procedure will be known as ‘stratified random sampling’. Other modes of selection of sampling such as systematic sampling can also be adopted within the different strata.

Estimation of mean and variance

We shall assume that the population of N units is first divided into k strata of $N_1, N_2, \ldots, N_k$ units respectively. These strata are non-overlapping and together they comprise the whole population, so that

$$N_1 + N_2 + \ldots + N_k = N. \quad (5.21)$$

When the strata have been determined, a sample is drawn from each stratum, the selection being made independently in each stratum. The sample sizes within the strata are denoted by $n_1, n_2, \ldots, n_k$ respectively, so that

$$n_1 + n_2 + \ldots + n_3 = n \quad (5.22)$$

Let y_{tj} ($j = 1, 2, \ldots, N_t$; $t = 1, 2, \ldots k$) be the value of the characteristic under study for the j the unit in the tth stratum. In this case, the population mean in the tth stratum is given by

$$\bar{Y}_t = \frac{1}{N_t}\sum_{j=1}^{N_t} y_{tj}, \ (t = 1, 2, \ldots, k) \quad (5.23)$$

The overall population mean is given by

$$\bar{Y} = \frac{1}{N}\sum_{t=1}^{k} N_t \bar{Y}_t \quad (5.24)$$

The estimate of the population mean $\bar{Y}$, in this case will be obtained by

$$\hat{\bar{Y}} = \frac{\sum_{t=1}^{k} N_t \bar{y}_t}{N} \quad (5.25)$$

where $$\bar{y}_t = \sum_{j=1}^{n_t} \frac{y_{tj}}{n_t} \quad (5.26)$$

Estimate of the variance of $\hat{\bar{Y}}$ is given by

$$\hat{V}(\hat{\bar{Y}}) = \frac{1}{N^2}\sum_{t=1}^{k} N_t(N_t - n_t)\frac{s_{t(y)}^2}{n_t} \quad (5.27)$$

where $$s_{t(y)}^2 = \sum_{j=1}^{n_t} \frac{(y_{tj} - \bar{y}_t)^2}{n_t - 1} \quad (5.28)$$

Stratification, if properly done as explained in the previous sections, will usually give lower variance for the estimated population total or mean than a simple random sample of the same size. However, a stratified sample taken without due care and planning may not be better than a simple random sample.

Numerical illustration of calculating the estimate of mean volume per hectare of a particular species and its standard error from a stratified random sample of compartments selected independently with equal probability in each stratum is given below.

A forest area consisting of 69 compartments was divided into three strata containing compartments 1-29, compartments 30-45, and compartments 46 to 69 and 10, 5 and 8 compartments respectively were chosen at random from the three strata. The serial numbers of the selected compartments in each stratum are given in column (4) of Table 5.3. The corresponding observed volume of the particular species in each selected compartment in m^3/ha is shown in column (5).

Table 5.3. Illustration of estimation of parameters under stratified sampling

Stratum number	Total number of units in the stratum (N_t)	Number of units sampled (n_t)	Selected sampling unit number	Volume (m^3/ha) (y_{t_j})	($y_{t_j}^2$)
(1)	(2)	(3)	(4)	(5)	(6)
			1	5.40	29.16
			18	4.87	23.72
			28	4.61	21.25
I			12	3.26	10.63

			20	4.96	24.60
			19	4.73	22.37
			9	4.39	19.27
			6	2.34	5.48
			17	4.74	22.47
			7	2.85	8.12
Total	29	10	..	42.15	187.07
			43	4.79	22.94
II			42	4.57	20.88
			36	4.89	23.91
			45	4.42	19.54
			39	3.44	11.83
Total	16	5	..	22.11	99.10
			59	7.41	54.91
			50	3.70	13.69
			49	5.45	29.70
III			58	7.01	49.14
			54	3.83	14.67
			69	5.25	27.56
			52	4.50	20.25
			47	6.51	42.38
Total	24	8	..	43.66	252.30

Step 1. Compute the following quantities.

N = (29 + 16 + 24) = 69

n = (10 + 5 + 8) = 23

$\bar{y}_t$ = 4.215, = 4.422, = 5.458

Step 2. Estimate of the population mean $\bar{Y}$ using Equation (3) is

$$\hat{\bar{Y}} = \frac{\sum_{t=1}^{3} N_t \bar{y}_t}{N} = \frac{(29 \times 4.215) + (16 \times 4.422) + (24 \times 5.458)}{69} = \frac{323.979}{69} = 4.70$$

Step 3. Estimate of the variance of $\hat{\bar{Y}}$ using Equation (5) as

$$\hat{V}(\hat{\bar{Y}}) = \frac{1}{N^2} \sum_{t=1}^{3} N_t \left(N_t - n_t\right) \frac{s_{t(y)}^2}{n_t}$$

In this example,

$$s_{1(y)}^2 = \frac{187.07 - \frac{(42.15)^2}{10}}{9} = \frac{9.41}{9} = 1.046$$

$$s_{2(y)}^2 = \frac{99.10 - \frac{(22.11)^2}{5}}{4} = \frac{1.33}{4} = 0.333$$

$$s_{3(y)}^2 = \frac{252.30 - \frac{(43.66)^2}{8}}{7} = \frac{14.03}{7} = 2.004$$

$$\hat{V}(\hat{\bar{Y}}) = \left(\frac{1}{69}\right)^2 \left[\left(\frac{29 \times 19}{10} \times 1.046\right) + \left(\frac{16 \times 11}{5} \times 0.333\right) + \left(\frac{24 \times 16}{8} \times 2.004\right)\right]$$

$$= \frac{165.5482}{4761} = 0.03477$$

$$SE(\hat{\bar{Y}}) = \sqrt{0.03477} = 0.1865$$

$$RSE(\hat{\bar{Y}}) = \frac{SE(\hat{\bar{Y}}) \times 100}{\hat{\bar{Y}}} \qquad (5.29)$$

$$= \frac{0.1865 \times 100}{4.70} = 3.97\%$$

Now, if we ignore the strata and assume that the same sample of size *n* = 23, formed a simple random sample from the population of *N* = 69, the estimate of the population mean would reduce to

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{42.15 + 22.11 + 43.66}{23} = \frac{107.92}{23} = 4.69$$

Estimate of the variance of the mean $\bar{y}$ is

$$\hat{V}(\bar{y}) = \frac{N-n}{Nn}s^2$$

where

$$s^2 = \frac{538.47 - \frac{(107.92)^2}{23}}{22}$$

$$= \frac{32.09}{22} = 1.4586$$

so that

$$\hat{V}(\bar{y}) = \frac{(69-23)}{69 \times 23} \times 1.4586$$

$$= \frac{2.9172}{69} = 0.04230$$

$$SE(\bar{y}) = \sqrt{0.04230} = 0.2057$$

$$RSE(\bar{y}) = \frac{0.2057 \times 100}{4.69} = 4.39\%$$

The gain in precision due to stratification is computed by

$$\frac{\hat{V}(\hat{\bar{Y}})_{srs}}{\hat{V}(\hat{\bar{Y}})_{st}} \times 100 = \frac{0.04230}{0.03477} \times 100$$

= 121.8

Thus the gain in precision is 21.8%.

Multistage sampling

With a view to reduce cost and/or to concentrate the field operations around selected points and at the same time obtain precise estimates, sampling is sometimes carried out in stages. The procedure of first selecting large sized units and then choosing a specified number of sub-units from the selected large units is known as sub-sampling. The large units are called ‘first stage units’ and the sub-units the ‘second stage units’. The procedure can be easily generalised to three stage or multistage samples. For example, the sampling of a forest area may be done in three stages, firstly by selecting a sample of compartments as first stage units, secondly, by choosing a sample of topographical sections in each selected compartment and lastly, by taking a number of sample plots of a specified size and shape in each selected topographical section.

The multistage sampling scheme has the advantage of concentrating the sample around several ‘sample points’ rather than spreading it over the entire area to be surveyed. This reduces considerably the cost of operations of the survey and helps to reduce the non-sampling errors by efficient supervision. Moreover, in forest surveys it often happens that detailed information may be easily available only for groups of sampling units but not for individual units. Thus, for example, a list of compartments with details of area may be available but the details of the topographical sections in each compartment may not be available. Hence if compartments are selected as first stage units, it may be practicable to collect details regarding the topographical sections for selected compartments only and thus use a two-stage sampling scheme without attempting to make a frame of the topographical sections in all compartments. The multistage sampling scheme, thus, enables one to use an incomplete sampling frame of all the sampling units and to properly utilise the information already available at every stage in an efficient manner.

The selection at each stage, in general may be either simple random or any other probability sampling method and the method may be different at the different stages. For example one may select a simple random sample of compartments and take a systematic line plot survey or strip survey with a random start in the selected compartments.

Two-stage simple random sampling

When at both stages the selection is by simple random sampling, method is known as two stage simple random sampling. For example, in estimating the weight of grass in a forest area, consisting of 40 compartments, the compartments may be considered as primary sampling units.

Out of these 40 compartments, n = 8 compartments may be selected randomly using simple random sampling procedure as illustrated in Section 5.2.1. A random sample of plots either equal or unequal in number may be selected from each selected compartment for the measurement of the quantity of grass through the procedure of selecting a simple random sample. It is then possible to develop estimates of either mean or total quantity of grass available in the forest area through appropriate formulae.

Parameter estimation under two-stage simple random sampling

Let the population consists of N first stage units and let M_i be the number of second stage units in the ith first stage unit. Let n first stage units be selected and from the ith selected first stage unit let m_i second stage units be chosen to form a sample of $m = \sum_{i=1}^{n} m_i$ units. Let y_{ij} be the value of the character for the jth second stage unit in the ith first stage unit.

An unbiased estimator of the population mean $\bar{Y} = \dfrac{\sum_{i}^{N}\sum_{j}^{M_i} y_{ij}}{\sum_{i=1}^{N} M_i}$ is obtained by Equation (5.30).

$$\hat{\bar{Y}} = \frac{1}{n\bar{M}}\sum_{i=1}^{n}\frac{M_i}{m_i}\sum_{j=1}^{m_i} y_{ij} \quad (5.30)$$

where $\bar{M} = \dfrac{\sum_{i=1}^{N} M_i}{N}$. (5.31)

The estimate of the variance of $\hat{\bar{Y}}$ is given by

$$\hat{V}(\hat{\bar{Y}}) = \left(\frac{1}{n} - \frac{1}{N}\right)s_b^2 + \frac{1}{nN}\sum_{i=1}^{n}\left(\frac{M_i}{\bar{M}}\right)^2\left(\frac{1}{m_i} - \frac{1}{M_i}\right)s_{w_i}^2 \quad (5.32)$$

where $s_b^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}\left(\dfrac{M_i}{\bar{M}}\bar{y}_i - \bar{y}\right)^2$ (5.33)

$$s_{w_i}^2 = \frac{1}{m_i - 1}\sum_{j=1}^{m_i}(y_{ij} - \bar{y}_i)^2 \quad (5.34)$$

The variance of $\hat{\hat{Y}}$ here can be noticed to be composed of two components. The first is a measure of variation between first stage units and the second, a measure of variation within first stage units. If $m_i = M_i$, the variance is given by the first component only. The second term, thus represents the contribution due to sub-sampling.

An example of the analysis of a two stage sample is given below. Table 5.4 gives data on weight of grass (mixed species) in kg from plots of size 0.025 ha selected from 8 compartments which were selected randomly out of 40 compartments from a forest area. The total forest area was 1800ha.

Table 5.4. Weight of grass in kg in plots selected through a two stage sampling procedure.

Plot	Compartment number								Total
	I	II	III	IV	V	VI	VII	VIII	
1	96	98	135	142	118	80	76	110	
2	100	142	88	130	95	73	62	125	
3	113	143	87	106	109	96	105	77	
4	112	84	108	96	147	113	125	62	
5	88	89	145	91	91	125	99	70	
6	139	90	129	88	125	68	64	98	
7	140	89	84	99	115	130	135	65	
8	143	94	96	140	132	76	78	97	
9	131	125	..	98	148	84	..	106	
10	..	116	..	..	..	105	..	..	
Total	1062	1070	872	990	1080	950	744	810	7578
m_i	9	10	8	9	9	10	8	9	72
Mean $(\bar{y}_i)$	118	107	109	110	120	95	93	90	842
M_i	1760	1975	1615	1785	1775	2050	1680	1865	14505
$s_{w_i}^2$	436.00	515.78	584.57	455.75	412.25	496.67	754.86	496.50	4152

$\frac{s_{w_i}^2}{m_i}$	48.44	51.578	73.07	50.63	45.80	49.667	94.35	55.167

Step1. Estimate the mean weight of grass in kg per plot using the formula in Equation (5.30).

$$\hat{\bar{Y}} = \frac{1}{n\bar{M}}\sum_{i=1}^{n}\frac{M_i}{m_i}\sum_{j=1}^{m_i}y_{ij}$$

$$\bar{M} = \frac{1}{N}\sum_{i=1}^{N} M_i = \frac{1}{40}\left(\frac{1800}{0.025}\right)$$

= 1800

Since $\sum M_i$ indicates the total number of second stage units, it can be obtained by dividing the total area (1800 ha) by the size of a second stage unit (0.025 ha).

Estimate of the population mean calculated using Equation (5.30) is

$$\hat{\bar{Y}} = \frac{1}{n\bar{M}}\sum_{i=1}^{n}\frac{M_i}{m_i}\sum_{j=1}^{m_i}y_{ij}$$

$= \frac{1523230}{(8)(40)}$ = 105.78

$$s_b^2 = \frac{1}{(8-1)}\left[\left(\frac{1760}{1800}\times 118 - 105.25\right)^2 + \left(\frac{1975}{1800}\times 107 - 105.25\right)^2 + \ldots + \left(\frac{1865}{1800}\times 90 - 105.25\right)^2\right]$$

=140.36

Estimate of variance of $\hat{\bar{Y}}$ obtained by Equation (5.32) is

$$\hat{V}(\hat{\bar{Y}}) = \left(\frac{1}{8} - \frac{1}{40}\right)140.3572 + \frac{1}{(8)(40)}(465.1024)$$

=15.4892

$SE(\hat{\bar{Y}}) = \sqrt{15.4892}$ = 3.9356

$$RSE(\hat{\bar{Y}}) = \frac{3.9356 \times 100}{105.78} = 3.72\%$$

Multiphase sampling

Multiphase sampling plays a vital role in forest surveys with its application extending over continuous forest inventory to estimation of growing stock through remote sensing. The essential idea in multiphase sampling is that of conducting separate sampling investigations in a sequence of phases starting with a large number of sampling units in the first phase and taking only a subset of the sampling units in each successive phase for measurement so as to estimate the parameter of interest with added precision at relatively lower cost utilizing the relation between characters measured at different phases. In order to keep things simple, further discussion in this section is restricted to only two phase sampling.

A sampling technique which involves sampling in just two phases (occasions) is known as two phase sampling. This technique is also referred to as double sampling. Double sampling is particularly useful in situations in which the enumeration of the character under study (main character) involves much cost or labour whereas an auxiliary character correlated with the main character can be easily observed. Thus it can be convenient and economical to take a large sample for the auxiliary variable in the first phase leading to precise estimates of the population total or mean of the auxiliary variable. In the second phase, a small sample, usually a sub-sample, is taken wherein both the main character and the auxiliary character may be observed and using the first phase sampling as supplementary information and utilising the ratio or regression estimates, precise estimates for the main character can be obtained. It may be also possible to increase the precision of the final estimates by including instead of one, a number of correlated auxiliary variables. For example, in estimating the volume of a stand, we may use diameter or girth of trees and height as auxiliary variables. In estimating the yield of tannin materials from bark of trees certain physical measurements like the girth, height, number of shoots, etc., can be taken as auxiliary variables.

Like many other kinds of sampling, double sampling is a technique useful in reducing the cost of enumerations and increasing the accuracy of the estimates. This technique can be used very advantageously in resurveys of forest areas. After an initial survey of an area, the estimate of growing stock at a subsequent, period, say 10 or 15 years later, and estimate of the change in growing stock can be obtained based on a relatively small sample using double sampling technique.

Another use of double sampling is in stratification of a population. A first stage sample for an auxiliary character may be used to sub-divide the population into strata in which the second (main) character varies little so that if the two characters are correlated, precise estimates of the main character can be obtained from a rather small second sample for the main character.

It may be mentioned that it is possible to couple with double sampling other methods of sampling like multistage sampling (sub-sampling)known for economy and enhancing the accuracy of the estimates. For example, in estimating the availability of grasses, canes, reeds, etc., a two-stage sample of compartments (or ranges) and topographical sections (or blocks) may be taken for the estimation of the effective area under the species and a sub-sample of topographical sections, blocks or plots may be taken for estimating the yield.

Selection of sampling units

In the simplest case of two phase sampling, simple random sampling can be employed in both the phases. In the first step, the population is divided into well identified sampling units and a sample is drawn as in the case of simple random sampling. The character *x* is measured on all the sampling units thus selected. Next, a sub-sample is taken from the already selected units using the method of simple random sampling and the main character of interest (*y*) is measured on the units selected. The whole procedure can also be executed in combination with other modes of sampling such as stratification or multistage sampling schemes.

Parameter estimation

Regression estimate in double sampling :

Let us assume that a random sample of *n* units has been taken from the population of *N* units at the initial phase to observe the auxiliary variable *x* and that a random sub-sample of size *m* is taken where both *x* and the main character *y* are observed.

Let $\bar{x}_{(n)}$ = mean of *x* in the first large sample = $\bar{x}_{(n)} = \sum_{i=1}^{n} \frac{x_i}{n}$ (5.35)

$\bar{x}_{(m)}$ = mean of *x* in the second sample = $\bar{x}_{(m)} = \sum_{i=1}^{m} \frac{x_i}{m}$ (5.36)

$\bar{y}$ = mean of y in the second sample = $\bar{y} = \sum_{i=1}^{m} \frac{y_i}{m}$ (5.37)

We may take $\bar{y}$ as an estimate of the population mean $\bar{Y}$. However utilising the previous information on the units sampled, a more precise estimate of $\bar{Y}$ can be obtained by calculating the regression of *y* on *x* and using the first sample as providing supplementary information. The regression estimate of $\bar{Y}$ is given by

$\bar{y}_{(drg)} = \bar{y} + b(\bar{x}_{(n)} - \bar{x}_{(m)})$ (5.38)

where the suffix *(drg)* denotes the regression estimate using double sampling and *b* is the regression coefficient of *y* on *x* computed from the units included in the second sample of size *m*. Thus

$$b = \frac{\sum_{i=1}^{m}\left(x_i - \bar{x}_{(m)}\right)\left(y_i - \bar{y}\right)}{\sum_{i=1}^{m}\left(x_i - \bar{x}_{(m)}\right)^2} \quad (5.39)$$

The variance of the estimate is approximately given by,

$$V(\bar{y})_{(drg)} = \frac{s_{y.x}^2}{m} + \frac{s_{y.x}^2 - s_y^2}{n} \quad (5.40)$$

where $$s_{y.x}^2 = \frac{1}{m-2}\left[\sum_{i=1}^{m}(y_i - \bar{y})^2 - b^2\sum_{i=1}^{m}(x_i - \bar{x}_{(m)})^2\right] \quad (5.41)$$

$$s_y^2 = \sum_{i=1}^{m}\frac{(y_i - \bar{y})^2}{m-1} \quad (5.42)$$

(ii) Ratio estimate in double sampling

Ratio estimate is used mainly when the intercept in the regression line between *y* and *x* is understood to be zero. The ratio estimate of the population mean $\bar{Y}$ is given by

$$\bar{y}_{(dra)} = \frac{\bar{y}}{\bar{x}_{(m)}}\bar{x}_{(n)} \quad (5.43)$$

where $\bar{y}_{dra}$ denotes the ratio estimate using double sampling. The variance of the estimate is approximately given by

$$V(\bar{y}_{dra}) = \frac{s_y^2 - 2\hat{R}s_{yx} + \hat{R}^2 s_x^2}{m} + \frac{2\hat{R}s_{yx} - \hat{R}^2 s_x^2}{n} \quad (5.44)$$

where

$$s_y^2 = \frac{\sum_{i=1}^{m}\left(y_i - \bar{y}\right)^2}{m-1} \quad (5.45)$$

$$s_{yx} = \frac{\sum_{i=1}^{m} (y_i - \bar{y})(x_i - \bar{x}_{(m)})}{m - 1} \quad (5.46)$$

$$s_x^2 = \frac{\sum_{i-1}^{m} (x_i - \bar{x}_m)^2}{m - 1} \quad (5.47)$$

$$\hat{R} = \frac{\bar{y}}{\bar{x}_{(m)}} \quad (5.48)$$

An example of analysis of data from double sampling using regression and ratio estimate is given below. Table 5.5 gives data on the number of clumps and the corresponding weight of grass in plots of size 0.025 ha, obtained from a random sub-sample of 40 plots taken from a preliminary sample of 200 plots where only the number of clumps was counted.

Table 5.5. Data on the number of clumps and weight of grass in plots selected through a two phase sampling procedure.

< TD WIDTH="16%" VALIGN="TOP">

60

Serial number	Number of clumps	Weight in kg	Serial number	Number of clumps	Weight in kg
	(*x*)	(*y*)		(*x*)	(*y*)
1	459	68	21	245	25
2	388	65	22	185	50
3	314	44	23	59	16
4	35	15	24	114	22
5	120	34	25	354	59
6	136	30	26	476	63
7	367	54	27	818	92
8	568	69	28	709	64

9	764	72	29	526	72
10	607	65	30	329	46
11	886	95	31	169	33
12	507	32	648	74	
13	417	72	33	446	61
14	389	60	34	86	32
15	258	50	35	191	35
16	214	30	36	342	40
17	674	70	37	227	40
18	395	57	38	462	66
19	260	45	39	592	68
20	281	36	40	402	55

Here, n = 200, m = 40. The mean number of clumps per plot as observed from the preliminary sample of 200 plots was $\bar{x}_{(n)}$ = 374.4.

$$\sum_{i=1}^{40} x_i = 15419, \quad \sum_{i=1}^{40} y_i = 2104,$$

$$\sum_{i=1}^{40} x_i^2 = 7744481, \quad \sum_{i=1}^{40} y_i^2 = 125346, \quad \sum_{i=1}^{40} x_i y_i = 960320$$

$$\sum_{1}^{40} (x_i - \bar{x}_{(m)})^2 = \sum_{i=1}^{40} x_i^2 - \frac{\left(\sum_{1}^{40} x_i\right)^2}{40} = 7744481 - \frac{\left(15419^2\right)}{40} = 1800842$$

$$\sum_{1}^{40} (y_i - \bar{y})^2 = \sum_{i=1}^{40} y_i^2 - \frac{\left(\sum_{i=1}^{40} y_i\right)^2}{40} = 125346 - \frac{(2104)^2}{40} = 14675.6$$

$$\sum_{1}^{40}\left(x_i - \bar{x}_{(m)}\right)\left(y_i - \bar{y}\right) = \sum_{1}^{40} x_i y_i - \frac{\sum_{1}^{40} x_i \sum_{1}^{40} y_i}{40} = 960320 - \frac{15419 \times 2104}{40} = 149280.6$$

Mean number of clumps per plot from the sub-sample of 40 plots is

$$\bar{x}_{(m)} = \frac{15419}{40} = 385.5$$

Mean weight of clumps per plot from the sub-sample of 40 plots

$$\bar{y} = \frac{2104}{40} = 52.6$$

The regression estimate of the mean weight of grass in kg per plot is obtained by using Equation (5.38) where the regression coefficient *b* obtained using Equation (5.39) is

$$b = \frac{149280.6}{1800842} = 0.08$$

Hence, $\bar{y}_{(drg)} = 52.6 + 0.08(374.4 - 385.5)$

= 52.6 - 0.89

= 51.7 kg /plot

$$s_{y.x}^2 = \frac{1}{40-2}\left[14675.6 - (0.08)^2(1800842)\right]$$

= 82.9

$$s_y^2 = \frac{14675.6}{39}$$

=376.297

The variance of the estimate is approximately given by Equation (5.40)

$$V(\bar{y})_{(drg)} = \frac{82.9}{40} + \frac{82.9 - 376.297}{200} \quad (5.40)$$

= 3.5395

The ratio estimate of the mean weight of grass in kg per plot is given by Equation (5.43)

$$\bar{y}_{(dra)} = \frac{52.6}{385.5}(374.4)$$

= 51.085

$$s_{yx} = \frac{149280.6}{40-1}$$

= 3827.708

$$s_x^2 = \frac{1800842}{40-1}$$

= 46175.436

$$\hat{R} = \frac{52.6}{385.5}$$

= 0.1364

The variance of the estimate is approximately given by Equation (5.44) is

$$V(\bar{y}_{dra}) = \frac{376.297 - 2(0.1364)(3827.708) + (0.1364)^2(46175.436)}{40}$$

$$+ \frac{(2)(0.1364)(3827.708) - (0.1364)^2(46175.436)}{200}$$

= 5.67

Probability Proportional to Size (PPS) sampling

In many instances, the sampling units vary considerably in size and simple random sampling may not be effective in such cases as it does not take into account the possible importance of the larger units in the population. In such cases, it has been found that ancillary information about the size of the units can be gainfully utilised in selecting the sample so as to get a more efficient estimator of the population parameters. One such method is to assign unequal probabilities for selection to different units of the population. For example, villages with larger geographical area are likely to have larger area under food crops and in estimating the production, it would be desirable to adopt a sampling scheme in which villages are selected with probability proportional to geographical area. When units vary in their size and the variable under study is directly related with the size of the unit, the probabilities may be assigned proportional to the size of the unit. This type of sampling where the probability of selection is proportion to the size of the unit is known as ‘PPS Sampling’. While sampling successive units from the population, the units already selected can be replaced back in the population or not. In the following, PPS sampling with replacement of sampling units is discussed as this scheme is simpler compared to the latter.

Methods of selecting a pps sample with replacement

The procedure of selecting the sample consists in associating with each unit a number or numbers equal to its size and selecting the unit corresponding to a number chosen at random from the totality of numbers associated. There are two methods of selection which are discussed below:

(i) Cumulative total method: Let the size of the ith unit be x_i, (i = 1, 2, …, N). We associate the numbers 1 to x_i with the first unit, the numbers (x_1+1) to (x_1+x_2) with the second unit and so on such that the total of the numbers so associated is $X = x_1 + x_2 + \ldots + x_N$. Then a random number r is chosen at random from 1 to X and the unit with which this number is associated is selected.

For example, a village has 8 orchards containing 50, 30, 25, 40, 26, 44, 20 and 35 trees respectively. A sample of 3 orchards has to be selected with replacement and with probability proportional to number of trees in the orchards. We prepare the following cumulative total table:

Serial number of the orchard	Size (x_i)	Cumulative size	Numbers associated
1	50	50	1 - 50
2	30	80	51 - 80
3	25	105	81 -105
4	40	145	106 -145
5	26	171	146 - 171
6	44	215	172 - 215
7	20	235	216 - 235
8	35	270	236 - 270

Now, we select three random numbers between 1 and 270. The random numbers selected are 200, 116 and 47. The units associated with these three numbers are 6^{th}, 4^{th}, and 1^{st} respectively. And hence, the sample so selected contains units with serial numbers, 1, 4 and 6.

(ii) Lahiri's Method: We have noticed that the cumulative total method involves writing down the successive cumulative totals which is time consuming and tedious, especially with large populations. Lahiri in 1951 suggested an alternative procedure which avoids the necessity of writing down the cumulative totals. Lahiri's method consists in selecting a pair of random numbers, say (i,j) such that 1 £ i £ N and 1£ j £ M; where M is the maximum of the sizes of the N units of the population. If j £ X_i, the ith unit is selected: otherwise, the pair of random number is rejected and another pair is chosen. For selecting a sample of n units, the procedure is to be repeated till n units are selected. This procedure leads to the required probabilities of selection.

For instance, to select a sample of 3 orchards from the population in the previous example in this section, by Lahiri's method by PPS with replacement, as N = 8, M = 50 and n = 3, we have to select three pairs of random numbers such that the first random number is less than or equal to 8 and the second random number is less than or equal to 50. Referring to the random number table, three pairs selected are (2, 23) (7,8) and (3, 30). As in the third pair $j > X_i$, a fresh pair has to be selected. The next pair of random numbers from the same table is (2, 18) and hence, the sample so selected consists of the units with serial numbers 2, 7 and 2. Since the sampling unit 2 gets repeated in the sample, the effective sample size is two in this case. In order to get an effective sample size of three, one may repeat the sampling procedure to get another distinct unit.

Estimation procedure

Let a sample of n units be drawn from a population consisting of N units by PPS with replacement. Further, let (y_i, p_i) be the value and the probability of selection of the ith unit of the sample, i = 1, 2, 3, …., n.

An unbiased estimator of population mean is given by

$$\hat{\bar{Y}} = \frac{1}{nN}\sum_{i=1}^{n}\frac{y_i}{p_i} \quad (5.49)$$

An estimator of the variance of above estimator is given by

$$\hat{V}\left(\hat{\bar{Y}}\right) = \frac{1}{n(n-1\_N^2}\left(\sum_{i=1}^{N}\left(\frac{y_i}{p_i}\right)^2 - n\hat{Y}^2\right) \quad (5.50)$$

where $p_i = \frac{x_i}{X}$, $\hat{Y} = N\hat{\bar{Y}}$

For illustration, consider the following example. A random sample 23 units out of 69 units were selected with probability proportional to size of the unit (compartment) from a forest area in U.P. The total area of 69 units was 14079 ha. The volume of timber determined for each selected compartment are given in Table 5.6 along with the area of the compartment.

Table 5. 6. Volume of timber and size of the sampling unit for a PPS sample of forest compartments.

Serialno.	Size in ha (x_i)	Relative size (x_i/X)	Volume in m^3 (y_i)	$\frac{y_i}{p_i} = v_i$	$(v_i)^2$
1	135	0.0096	608	63407.644	4020529373.993
2	368	0.0261	3263	124836.351	15584114417.014
3	374	0.0266	877	33014.126	1089932493.652
4	303	0.0215	1824	84752.792	7183035765.221
5	198	0.0141	819	58235.864	3391415813.473
6	152	0.0108	495	45849.375	2102165187.891
7	264	0.0188	1249	66608.602	4436705896.726
8	235	0.0167	1093	65482.328	4287935235.716
9	467	0.0332	1432	43171 .580	1863785345.581
10	458	0.0325	3045	93603.832	8761677342.194
11	144	0.0102	410	40086.042	1606890736.502
12	210	0.0149	1460	97882.571	9580997789.469
13	467	0.0332	1432	43171.580	1863785345.581
14	458	0.0325	3045	93603.832	8761677342.194
15	184	0.0131	1003	76745.853	5889925992.739
16	174	0.0124	834	67482.103	4553834285.804
17	184	0.0131	1003	76745.853	5889925992.739
18	285	0.0202	2852	140888.800	19849653965.440
19	621	0.0441	4528	102656.541	10538365422.979
20	111	0.0079	632	80161.514	6425868248.777
21	374	0.0266	877	33014.126	1089932493.652

22	64	0.0045	589	129570.797	16788591402.823
23	516	0.0367	1553	42373.424	1795507096.959
				1703345.530	147356252987.120

Total area X = 14079 ha.

An unbiased estimator of population mean is obtained by using Equation (5.49).

$$\hat{\bar{Y}} = \frac{1}{(23)(69)}(1703345.530)$$

= 1073.312

An estimate of the variance of $\hat{\bar{Y}}$ is obtained through Equation (5.50).

$$\hat{V}\left(\hat{\bar{Y}}\right) = \frac{1}{23(23-1)(69)^2}\left(147356252987.120 - (23)(67618.632)\right)$$

= 17514.6

And the standard error of $\bar{Y}$ is $\sqrt{17514.6}$ = 132.343.

--------D Guha-------

Appendix K

The F Distribution

The F distribution is an asymmetric distribution that has a minimum value of 0, but no maximum value. The curve reaches a peak not far to the right of 0, and then gradually approaches the horizontal axis the larger the F value is. The F distribution approaches, but never quite touches the horizontal axis.

The F distribution has two degrees of freedom, d_1 for the numerator, d_2 for the denominator. For each combination of these degrees of freedom there is a different F distribution. The F distribution is most spread out when the degrees of freedom are small. As the degrees of freedom increase, the F distribution the F distribution is less dispersed.

Figure 1.1 shows the shape of the distribution. The F value is on the horizontal axis, with the probability for each F value being represented by the vertical axis. The shaded area in the diagram represents the level of significance α shown in the table.

There is a different F distribution for each combination of the degrees of freedom of the numerator and denominator. Since there are so many F distributions, the F tables are organized somewhat differently than the tables for the other distributions. The three tables which follow are organized by the level of significance. The first table gives F values for that are associated with $\alpha = 0.10$ of the area in the right tail of the distribution. The second table gives the F values for $\alpha = 0.05$ of the area in the right tail, and the third table gives F values for the $\alpha = 0.01$ level of significance. In each of these tables, the F values are given for various combinations of degrees of freedom.

In order to use the F table, first select the significance level to be used,

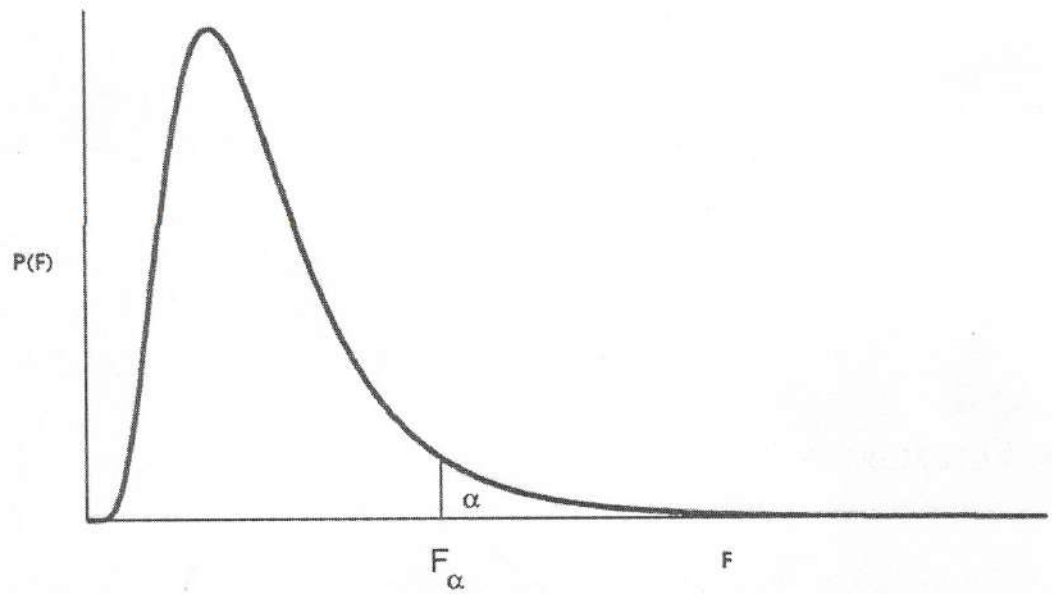


Figure K.1: The F distribution

Figure K.1: The F distribution

and then determine the appropriate combination of degrees of freedom. For example, if the $\alpha = 0.10$ level of significance is selected, use the first F table. If there are 5 degrees of freedom in the numerator, and 7 degrees of freedom in the denominator, the F value from the table is 2.88. This means that there is exactly 0.10 of the area under the F curve that lies to the right of $F = 2.88$.

When the significance level is $\alpha = 0.05$, use the second F table. If there are 20 degrees of freedom in the numerator, and 5 degrees of freedom in the denominator, then the critical F value is 4.56. This could be written

$$F_{20,5;0.05} = 4.56$$

That is, for 20 and 5 degrees of freedom, the F value that leaves exactly 0.05 of the area under the F curve in the right tail of the distribution is 4.56.

For the $\alpha = 0.01$ level of significance, the third F table is used. Suppose that there is 1 degree of freedom in the numerator and 12 degrees of freedom in the denominator. Then

$$F_{1,12;0.01} = 9.33.$$

An F value of 9.33 leaves exactly 0.01 of area under the curve in the right tail of the distribution when there are 1 and 12 degrees of freedom.

F Values for $\alpha = 0.10$

d_2 \ d_1	1	2	3	4	5	6	7	8	9
1	39.86	49.5	53.59	55.83	57.24	58.2	58.91	59.44	59.86
2	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38
3	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24
4	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94
5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32
6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96
7	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72
8	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56
9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.3	2.27
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96
21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93
23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88
27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87
28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87
29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86
30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74
120	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68
inf	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63

F Value for $\alpha = 0.10$

d_2 \ d_1	10	12	15	20	24	30	40	60	120	inf
1	60.19	60.71	61.22	61.74	62	62.26	62.53	62.79	63.06	63.33
2	9.39	9.41	9.42	9.44	9.45	9.46	9.47	9.47	9.48	9.49
3	5.23	5.22	5.20	5.18	5.18	5.17	5.16	5.15	5.14	5.13
4	3.92	3.90	3.87	3.84	3.83	3.82	3.80	3.79	3.78	3.76
5	3.30	3.27	3.24	3.21	3.19	3.17	3.16	3.14	3.12	3.10
6	2.94	2.90	2.87	2.84	2.82	2.80	2.78	2.76	2.74	2.72
7	2.70	2.67	2.63	2.59	2.58	2.56	2.54	2.51	2.49	2.47
8	2.54	2.50	2.46	2.42	2.40	2.38	2.36	2.34	2.32	2.29
9	2.42	2.38	2.34	2.30	2.28	2.25	2.23	2.21	2.18	2.16
10	2.32	2.28	2.24	2.20	2.18	2.16	2.13	2.11	2.08	2.06
11	2.25	2.21	2.17	2.12	2.10	2.08	2.05	2.03	2.00	1.97
12	2.19	2.15	2.10	2.06	2.04	2.01	1.99	1.96	1.93	1.90
13	2.40	2.10	2.05	2.01	1.98	1.96	1.93	1.90	1.88	1.85
14	2.10	2.05	2.01	1.96	1.94	1.91	1.89	1.86	1.83	1.80
15	2.06	2.02	1.97	1.92	1.90	1.87	1.85	1.82	1.79	1.76
16	2.03	1.99	1.94	1.89	1.87	1.84	1.81	1.78	1.75	1.72
17	2.00	1.96	1.91	1.86	1.84	1.81	1.78	1.75	1.72	1.69
18	1.98	1.93	1.89	1.84	1.81	1.78	1.75	1.72	1.69	1.66
19	1.96	1.91	1.86	1.81	1.79	1.76	1.73	1.70	1.67	1.63
20	1.94	1.89	1.84	1.79	1.77	1.74	1.71	1.68	1.64	1.61
21	1.92	1.87	1.83	1.78	1.75	1.72	1.69	1.66	1.62	1.59
22	1.90	1.86	1.81	1.76	1.73	1.70	1.67	1.64	1.60	1.57
23	1.89	1.84	1.80	1.74	1.72	1.69	1.66	1.62	1.59	1.55
24	1.88	1.83	1.78	1.73	1.70	1.67	1.64	1.61	1.57	1.53
25	1.87	1.82	1.77	1.72	1.69	1.66	1.63	1.59	1.56	1.52
26	1.86	1.81	1.76	1.71	1.80	1.65	1.61	1.58	1.54	1.50
27	1.85	1.80	1.75	1.70	1.67	1.64	1.60	1.57	1.53	1.49
28	1.84	1.79	1.74	1.69	1.66	1.63	1.59	1.56	1.52	1.48
29	1.83	1.78	1.73	1.68	1.65	1.62	1.58	1.55	1.51	1.47
30	1.82	1.77	1.72	1.67	1.64	1.61	1.57	1.54	1.50	1.46
40	1.76	1.71	1.66	1.61	1.57	1.54	1.51	1.47	1.42	1.38
60	1.71	1.66	1.60	1.54	1.51	1.48	1.44	1.40	1.35	1.29
120	1.65	1.60	1.55	1.48	1.45	1.41	1.37	1.32	1.26	1.19
inf	1.60	1.55	1.49	1.42	1.38	1.34	1.30	1.24	1.17	1.00

F Values for $\alpha = 0.05$

	d_1								
d_2	1	2	3	4	5	6	7	8	9
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5
2	18.51	19.00	19.16	19.25	19.3	19.33	19.35	19.37	19.38
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96
inf	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88

F Values for $\alpha = 0.05$

d_2 \ d_1	10	12	15	20	24	30	40	60	120	inf
1	241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3
2	19.4	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.5
3	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	1.91	1.83	1.75	1.66	1.10	1.55	1.50	1.43	1.35	1.25
inf	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

F Values for $\alpha = 0.01$

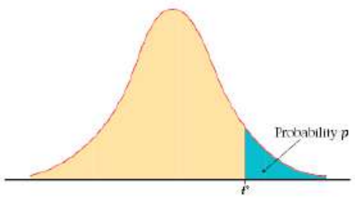
d_2 \ d_1	1	2	3	4	5	6	7	8	9
1	4052	4999.5	5403	5625	5764	5859	5928	5982	6022
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35
10	10.04	7.56	6.55	5.99	5.64	5.39	5.2	5.06	4.94
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.14
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56
inf	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41

F Values for $\alpha = 0.01$

d_2 \ d_1	10	12	15	20	24	30	40	60	120	inf
1	6056	6106	6157	6209	6235	6261	6287	6313	6339	6366
2	99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.49	99.50
3	27.23	27.05	26.87	26.69	26.60	26.50	26.41	26.32	26.22	26.13
4	14.55	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46
5	10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
7	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
12	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17
14	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
15	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
16	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75
17	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65
18	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49
20	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
21	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17
26	3.09	2.96	2.81	2.66	2.58	2.50	2.42	2.33	2.23	2.13
27	3.06	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	2.10
28	3.03	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06
29	3.00	2.87	2.73	2.57	2.49	2.41	2.33	2.23	2.14	2.03
30	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
40	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80
60	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38
inf	2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00

t-distribution table

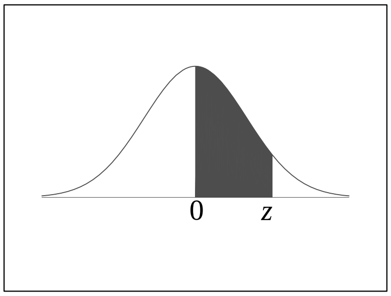
Areas in the upper tail are given along the top of the table. Critical t* values are given in the table.



df	0.1	0.05	0.025	0.02	0.01	0.005
1	3.078	6.314	12.706	15.895	31.821	63.657
2	1.886	2.920	4.303	4.849	6.965	9.925
3	1.638	2.353	3.182	3.482	4.541	5.841
4	1.533	2.132	2.776	2.999	3.747	4.604
5	1.476	2.015	2.571	2.757	3.365	4.032
6	1.440	1.943	2.447	2.612	3.143	3.707
7	1.415	1.895	2.365	2.517	2.998	3.499
8	1.397	1.860	2.306	2.449	2.896	3.355
9	1.383	1.833	2.262	2.398	2.821	3.250
10	1.372	1.812	2.228	2.359	2.764	3.169
11	1.363	1.796	2.201	2.328	2.718	3.106
12	1.356	1.782	2.179	2.303	2.681	3.055
13	1.350	1.771	2.160	2.282	2.650	3.012
14	1.345	1.761	2.145	2.264	2.624	2.977
15	1.341	1.753	2.131	2.249	2.602	2.947
16	1.337	1.746	2.120	2.235	2.583	2.921
17	1.333	1.740	2.110	2.224	2.567	2.898
18	1.330	1.734	2.101	2.214	2.552	2.878
19	1.328	1.729	2.093	2.205	2.539	2.861
20	1.325	1.725	2.086	2.197	2.528	2.845
21	1.323	1.721	2.080	2.189	2.518	2.831
22	1.321	1.717	2.074	2.183	2.508	2.819
23	1.319	1.714	2.069	2.177	2.500	2.807
24	1.318	1.711	2.064	2.172	2.492	2.797
25	1.316	1.708	2.060	2.167	2.485	2.787
26	1.315	1.706	2.056	2.162	2.479	2.779
27	1.314	1.703	2.052	2.158	2.473	2.771
28	1.313	1.701	2.048	2.154	2.467	2.763
29	1.311	1.699	2.045	2.150	2.462	2.756
30	1.310	1.697	2.042	2.147	2.457	2.750
31	1.309	1.696	2.040	2.144	2.453	2.744
32	1.309	1.694	2.037	2.141	2.449	2.738
33	1.308	1.692	2.035	2.138	2.445	2.733
34	1.307	1.691	2.032	2.136	2.441	2.728
35	1.306	1.690	2.030	2.133	2.438	2.724
36	1.306	1.688	2.028	2.131	2.434	2.719
37	1.305	1.687	2.026	2.129	2.431	2.715
38	1.304	1.686	2.024	2.127	2.429	2.712
39	1.304	1.685	2.023	2.125	2.426	2.708
40	1.303	1.684	2.021	2.123	2.423	2.704
41	1.303	1.683	2.020	2.121	2.421	2.701
42	1.302	1.682	2.018	2.120	2.418	2.698
43	1.302	1.681	2.017	2.118	2.416	2.695
44	1.301	1.680	2.015	2.116	2.414	2.692
45	1.301	1.679	2.014	2.115	2.412	2.690
46	1.300	1.679	2.013	2.114	2.410	2.687
47	1.300	1.678	2.012	2.112	2.408	2.685
48	1.299	1.677	2.011	2.111	2.407	2.682
49	1.299	1.677	2.010	2.110	2.405	2.680
50	1.299	1.676	2.009	2.109	2.403	2.678

df	0.1	0.05	0.025	0.02	0.01	0.005
51	1.298	1.675	2.008	2.108	2.402	2.676
52	1.298	1.675	2.007	2.107	2.400	2.674
53	1.298	1.674	2.006	2.106	2.399	2.672
54	1.297	1.674	2.005	2.105	2.397	2.670
55	1.297	1.673	2.004	2.104	2.396	2.668
56	1.297	1.673	2.003	2.103	2.395	2.667
57	1.297	1.672	2.002	2.102	2.394	2.665
58	1.296	1.672	2.002	2.101	2.392	2.663
59	1.296	1.671	2.001	2.100	2.391	2.662
60	1.296	1.671	2.000	2.099	2.390	2.660
61	1.296	1.670	2.000	2.099	2.389	2.659
62	1.295	1.670	1.999	2.098	2.388	2.657
63	1.295	1.669	1.998	2.097	2.387	2.656
64	1.295	1.669	1.998	2.096	2.386	2.655
65	1.295	1.669	1.997	2.096	2.385	2.654
66	1.295	1.668	1.997	2.095	2.384	2.652
67	1.294	1.668	1.996	2.095	2.383	2.651
68	1.294	1.668	1.995	2.094	2.382	2.650
69	1.294	1.667	1.995	2.093	2.382	2.649
70	1.294	1.667	1.994	2.093	2.381	2.648
71	1.294	1.667	1.994	2.092	2.380	2.647
72	1.293	1.666	1.993	2.092	2.379	2.646
73	1.293	1.666	1.993	2.091	2.379	2.645
74	1.293	1.666	1.993	2.091	2.378	2.644
75	1.293	1.665	1.992	2.090	2.377	2.643
76	1.293	1.665	1.992	2.090	2.376	2.642
77	1.293	1.665	1.991	2.089	2.376	2.641
78	1.292	1.665	1.991	2.089	2.375	2.640
79	1.292	1.664	1.990	2.088	2.374	2.640
80	1.292	1.664	1.990	2.088	2.374	2.639
81	1.292	1.664	1.990	2.087	2.373	2.638
82	1.292	1.664	1.989	2.087	2.373	2.637
83	1.292	1.663	1.989	2.087	2.372	2.636
84	1.292	1.663	1.989	2.086	2.372	2.636
85	1.292	1.663	1.988	2.086	2.371	2.635
86	1.291	1.663	1.988	2.085	2.370	2.634
87	1.291	1.663	1.988	2.085	2.370	2.634
88	1.291	1.662	1.987	2.085	2.369	2.633
89	1.291	1.662	1.987	2.084	2.369	2.632
90	1.291	1.662	1.987	2.084	2.368	2.632
91	1.291	1.662	1.986	2.084	2.368	2.631
92	1.291	1.662	1.986	2.083	2.368	2.630
93	1.291	1.661	1.986	2.083	2.367	2.630
94	1.291	1.661	1.986	2.083	2.367	2.629
95	1.291	1.661	1.985	2.082	2.366	2.629
96	1.290	1.661	1.985	2.082	2.366	2.628
97	1.290	1.661	1.985	2.082	2.365	2.627
98	1.290	1.661	1.984	2.081	2.365	2.627
99	1.290	1.660	1.984	2.081	2.365	2.626
100	1.290	1.660	1.984	2.081	2.364	2.626

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	.4332	.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